Universidade de Brasília
Departamento de Estatística

# Comparações de modelos de regressão com resposta binária em delineamentos transversais

**Alyne Stuckert Florentino**

Trabalho de conclusão de curso do primeiro semestre de 2017 do Departamento de Estatística da Universidade de Brasília.

**Brasília**

**2017**

Alyne Stuckert Florentino 120006235

# Comparações de modelos de regressão com resposta binária em delineamentos transversais

Relatório final

Prof. **Eduardo Freitas da Silva**

Trabalho de conclusão de curso do primeiro semestre de 2017 do Departamento de Estatística da Universidade de Brasília.

**Brasília**

**2017**

**Resumo**

O interesse do presente estudo é uma comparação entre modelos de regressão para resposta binária, essas modelagens possibilitam a estimação da razão de prevalência (RP) que é muito utilizada em pesquisas epdemiológicas pois, além de ser de fácil interpretação, consegue calcular a prevalência de determinado desfecho entre dois grupos de interesse. As modelagens comparadas são: regressão logística, regressão de Poisson e regressão log-binomial. Os dois primeiros modelos não estimam diretamente a RP, tal que a regressão logística estima a razão de chances (odds ratio), ela se assemelha à RP em alguns casos de prevalência da doença. Já a regressão de Poisson se aproxima melhor pois, sua estimação é robusta permitindo melhores estimativas pontuais e intervalares, ela se aproxima por usar como distribuição base a distribuição de Poisson que pode ser admitida pois existe uma relação entre a distribuição binomial e a mesma. Por fim, o último modelo estima diretamente a RP porém tem algumas dificuldades de convergência, o que pode ser solucionado a partir de modificações nos parâmetros dos algoritimos ultilizados para a convergência. Essas modelagens calcularão as estimativas de RP e OR para o desfecho depressão pós-parto em mães interndas na UTI neonatal conciliada a variáveis de três blocos que estão encadeados e correlacionados de forma distal, intermediária e proximal a esse desfecho, formando assim uma estrutura hierárquica. Assim, o objetivo do estudo é verificar as diferenças das estimações de prevalência para essas três modelagens e verificar aquela que mais condiz com as situações definidas pelos dados usados.


**Palavras-chave:** Risco Relativo, Razão de Chances, Regressão Logística, Regressão de Poisson, Regressão Log-Binomial.

# Lista de Tabelas

# Sumário

# 1 Introdução

O presente trabalho pretende explicar algumas modelagens que abordam de forma individual a estimação de um determinado parâmetro de interesse em estudos epidemiológicos. Segundo Lima-Costa e Barreto (2003, p.191),

> A Epidemiologia é definida como o estudo da distribuição e dos determinantes das doenças ou condições relacionadas à saúde em populações especificadas. Mais recentemente, foi incorporada à definição de Epidemiologia a "aplicação desses estudos para controlar problemas de saúde".

Esses estudos têm como finalidade a verificação de uma causa ou relações entre doenças e possíveis variáveis que podem contribuir para a explicação das mesmas. Como dito, se especifica uma população com o evento de interesse, ou seja, a doença que se quer investigar, e, a partir das hipóteses, são determinadas variáveis de relação direta ou indireta com a variável principal. Em estudos epidemiológicos, algumas dessas variáveis são tratadas com outras nomenclaturas, como o desfecho que é a variável resposta do estudo. As variáveis explicativas são classificadas como fator de risco ou de confusão. O fator de risco está diretamente associado à variável resposta, em que geralmente é a variável de interesse em se testar a associação. As variáveis de confusão são secundárias na associação, ou seja, estão relacionadas com a resposta e o fator de risco e precisam ser controladas. Geralmente essas são variáveis socioeconômicas e sóciodemográficas. Usualmente, as variáveis de interesse são qualitativas nominais; mais que isso, são variáveis dicotômicas. Pode-se defini-las, de acordo com Bragança (1977), como variáveis que aparecem naturalmente ou pela discretização de uma variável contínua ou pela junção dos níveis de uma variável politômica. Em geral elas expressão o acontecimento ou não de um evento de interesse e se convencionou a denotação para cada observação i como a resposta $y_i = 1$, se o evento ocorrer, e $y_i = 0$, caso contrário.

Este estudo se caracteriza por conter esse tipo de variável. No caso, elas já são dicotômicas naturalmente, pois, de acordo com a análise, deseja-se investigar uma certa doença que possui dois níveis de classificação, ou seja, as respostas sim ou não.

Alguns planos de desenvolvimento de um estudo levam em consideração uma estrutura experimental ou observacional. Neste, a finalidade é a extração das informações pela observação dos dados coletados, em que o pesquisador não intervem na obtenção dos resultados, caracterizando assim um estudo observacional. Com relação ao tempo especificado para os dados, apenas obteve-se um retrato exato de determinada população em um tempo específico, como sendo uma fotografia no tempo, o que caracteriza um estudo transversal. A maior vantagem desse tipo de estudo é a obtenção rápida dos resultados, já que o tempo contabilizado corresponde apenas ao da coleta e análise dos dados, o que é esperado ser menor do que quando se tem um estudo que deseja acompanhar um grupo da população observando características durante alguns anos, por exemplo, estes chamados de estudos de coorte. Ele também se caracteriza por ter um baixo custo, ser mais acessível em termos logísticos e torna-se não sensível à perda de indivíduos, por não se tratar de um estudo com período de seguimento. Os estudos transversais também podem se tornar bases para análises mais profundas, pela suas facilidades, podem ser feitos para um pré-estudo e assim revelar possíveis associações de interesse, que colaborarão para finalidades futuras. Entretanto, como todo estudo, tem-se também suas desvantagens, como por exemplo, a ausência de acompanhamento temporal dos indivíduos, a impossibilidade de afirmar uma relação de causa e efeito das variáveis e também a exigência de uma quantidade suficientemente grande de amostra, o que não é compatível com uma pesquisa para doenças raras por exemplo.

Tradicionalmente o modelo de regressão logística é empregado neste caso, com as referidas características citadas. O mesmo não é linear e precisa de uma função que o linearize para assim ser capaz de fornecer interpretações condizentes com a hipótese estipulada. O trabalho tratará detalhadamente da teoria para a formulação desse modelo, tanto no âmbito simples como no múltiplo, além de determinar cada particularidade quando se têm diferentes tipos de variáveis. A consequência será a possibilidade da obtenção de uma estimativa similar para uma principal estatística utilizada na epidemiologia, a razão de prevalência. Essa medida é apta em traduzir a associação entre algumas características, ou seja, variáveis explicativas com relação à variável resposta. Portanto, a principal estatística para esses estudos é essa medida. Geralmente é estimada por uma divisão de probabilidades, ou seja, a probabilidade de ter a doença sob não ter a doença, por exemplo, sendo essa probabilidade dada pela divisão da frequência de casos de interesse sob o número total de casos (espaço amostral) representada pela equação (1) abaixo.

$$RP = \frac{\pi_1}{\pi_2} \tag{1}$$

Também será avaliada esse modelagem logística e poderá se concluir o porquê de a mesma estar se tornando inutilizada para estimar a razão de prevalência. Novas modelagens foram testadas e explanadas teoricamente proporcionando resultados na maioria dos casos melhores do que aqueles fornecidos pela regressão logística.

Atualmente outras duas modelagens estão sendo comumente utilizadas para este tipo de estudo e para estimar a prevalência. São elas: a regressão de Poisson com variância robusta e a regressão log-binomial. Como acontece na regressão logística, os outros dois tipos de regressão utilizados não são lineares e precisam de uma função que as linearize para ser possível fazer as devidas análises. Além disso, essas novas modelagens lidam melhor com a maioria dos casos na epdemiologia, usados principalmente em estudos de prevalência.

Como uma aplicação da teoria explicada no trabalho, também serão analisados dados reais de mulheres grávidas em que se deseja medir a depressão pós-parto. O banco de dados foi disponibilizado pela professora Karina Nascimento Costa do curso de medicina, da área de pediatria. As mães pesquisadas são questionadas sobre eventos traumáticos que ocorreram durante suas vidas, experiências obtidas durante e após o parto e algumas características sócio-demográficas como idade, escolaridade, estado civil, entre outras. O estudo visa compreender quais dessas situações podem contribuir para a variável de interesse, o desfecho e observar as discrepâncias entre as estimações pelas modelagens citadas. Deverá levar-se em consideração uma estrutura que relaciona esses grupos de variáveis hierarquicamente, logo também será incluído nos modelos de interesse esse aspecto.

# 2 Referencial Teórico

## 2.1 Modelo de Regrassão Logística

A literatura clássica oferece este tipo de modelagem para analisar estudos transversais e aqueles que possuem respostas de caráter binário. Ao se deparar com esse tipo de observação, em primeira instância deseja-se avaliar uma relação entre as variáveis, ou seja, os fatores de risco, e a variável resposta, o desfecho. Assim, pode-se averiguar essa correspondência por um gráfico de dispersão, que mostrará uma relação em que não é de fácil vizualização. Então uma possível solução inicial é a categorização da variável explicativa e, em seguida, o cálculo da proporção de respostas de interesse para cada classe da variável explicativa.

Um exemplo clássico usado por Lemeshow (2008), analisa um estudo sobre uma doença cardíaca, cuja desfecho é ter ou não a doença de coração ($y_i = 1$ ou $y_i = 0$). A variável explicativa usada é representada pela idade dos pacientes (fator de risco). Para averiguar uma possível associação linear, o autor define classes para a idade e assim a porporção de doentes para cada nível da idade, ficando imediato observar uma relação linear do crescimento da proporção de doentes com relação ao ponto médio de cada classe quando a idade tende a aumentar.

Infelizmente, quando se tem essa possível análise, perde-se informação dos dados, eles tornam-se resumidos dentro de cada classe. Logo, faz-se necessário o uso da modelagem proposta nesta seção.

O valor esperado para o desfecho na regressão logística será a probabilidade do índivíduo possuir a característica, ou seja, $Y_i = 1$ dado um fator de risco X que pode ser quantitativo contínuo ou discreto e qualitativo nominal ou ordinal com duas ou mais classes. Portanto essa probabilidade é representada por $\pi(x) = E(Y|X = x)$ e $0 \leq \pi(x) \leq 1$.

O modelo simples é representado por:

$$\pi(x) = \frac{exp(\beta_0 + \beta_1 x)}{1 + exp(\beta_0 + \beta_1 x)} \tag{2}$$

Pela equação (2) pode-se inferir que ela não linear e não é possível averiguar uma relação direta entre o desfecho e suas possíveis variáveis explicativas. Logo, usa-se uma função de ligação, chamada de logito, que é capaz de linearizar o modelo e torná-lo de fácil interpretação, como é no caso de uma regressão linear. O logito é denominado por g(x) e é representado por:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{3}$$

O modelo também pode ser escrito como:

$$y = \pi(x) + \varepsilon \tag{4}$$

Em que $y$ representa o desfecho e é igual à proporção de interesse dado um valor de $x$ (variável explicativa) mais um erro, que, por sua vez, possui distribuição binomial com parâmetros n e $\pi(x)$, além de serem independentes e identicamente distribuídos com média 0 e variância $\pi(x)(1-\pi(x))$. Essa então será a distribuição que irá embasar toda a análise.

A estimação para os parâmetros do modelo é feita pelo método da máxima verossimilhança, que é capaz de encontrar os parâmetros ideais para uma determinada distribuição e assim traduzir da melhor forma a amostra obtida. A função que pode ser associada, a partir de um erro com distribuição binomial, é dada por:

$$f(x,y) = \pi(x_i)^{y_i}[1-\pi(x_i)]^{n-y_i} \tag{5}$$

Os próximos passos afim de encontrar uma estimação para os parâmetros estão exemplificados pelas equações (6),(7) e (8). Determina-se a função de verossimilhança para a distribuição citada acima, aplica-se o logarítimo sob a nova função e maximiza-se esse logaritimo da função com relação ao parâmetro de interesse, igualando a zero. Assim pode-se encontrar a melhor e possível estimação com variância mínima, além de concordar com os critérios de um bom estimador.

A função de verossimilhança será:

$$L(\beta) = \prod \pi(x_i)^{y_i}[1-\pi(x_i)]^{n-y_i} \tag{6}$$

$$\frac{d}{d\beta_0} ln[L(\beta)] = 0 \tag{7}$$

$$\frac{d}{d\beta_1} ln[L(\beta)] = 0 \tag{8}$$

Assumindo $\sum y_i = Zi$, as equações chamadas de equações de verossimilhança são:

$$\sum [Z_i - n\pi(x_i)] = 0 \tag{9}$$

$$\sum [Z_i x_i - n x_i \pi(x_i)] = 0 \tag{10}$$

Após os passos citados e a substituição da função $\pi(x_i)$, é possível verificar que serão necessárias ferramentas mais aprimoradas que possam definir estas estimações, pois as expressões (9) e (10) não são lineares para seus respectivos parâmetros. É necessário o uso de cálculos numéricos, como o método de Newton-Raphson, em que auxiliará na definição das melhores estimativas para os parâmetros $\hat{\beta}_0$ e $\hat{\beta}_1$. É fácil adquirir essas estimações em softwares estatísticos.

Lemeshow (2008), também propõe outros métodos de estimação, *noniterative weighted least squares* e *discriminant function analysis*. Esses métodos alternativos são sensíveis com relação à normalidade assumida no pressuposto quando se tem um modelo com vários tipos de variáveis, por exemplo. Assim as estimações se tornam superestimadas. Portanto a melhor técnica usada é a de máxima verossimilhança.

Para a verificação da significância desses parâmetros são usados os testes da Razão da Verossimilhança, Wald e Score, sendo o mais utilizado o teste de Wald, que compara a estimação por maxima verossimilhança do parâmetro e seu desvio padrão. Suas hipóteses são $H_0 : \beta_1 = 0$ e $H_1 : \beta_1 \neq 0$ e sua estatística teste W segue uma distribuição normal padrão.

$$W = \frac{\hat{\beta}_i}{\hat{se}(\hat{\beta}_i)} \tag{11}$$

Os intervalos de confiança para os parâmetros também são baseados na estatística de Wald em que o desvio padrão, $\hat{se}(\beta_1)$, é assintótico. São dados por:

$$\hat{\beta}_0 \pm z_{1-\frac{\alpha}{2}} \hat{se}(\hat{\beta}_0) \quad \text{e} \quad \hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \hat{se}(\hat{\beta}_1) \tag{12}$$

### 2.1.1 Regressão logística múltipla

Usualmente é necessária uma melhor explicação do modelo e isso se dá por meio da inclusão de novas variáveis que estão relacionadas moderadamente ou fortemente com o desfecho. Por isso deve-se usar um modelo que estime mais parâmetros que representem por exemplo os fatores de risco variáveis de confusão. Este é chamado de modelo múltiplo e pode conter $p$ parâmetros para as $p-1$ possíveis variáveis que auxiliarão na explicação modelo. Assim existem mais de dois parâmetros a serem estimados, no caso $p$ parâmetros e para definir melhor como estão disposostas as variáveis do modelo, será feito o uso de notações matriciais, as quais facilitam o entendimento que parte do mesmo princípio do modelo simples.

$$\beta_{px1} = \begin{bmatrix} \beta_0 \\ \beta 1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad X_{i,px1} = \begin{bmatrix} 1 \\ X_{i1} \\ \vdots \\ X_{ip-1} \end{bmatrix}$$

O modelo para a Regressão Logística Múltipla tem a seguinte construção, o que não diferencia tanto do modelo simples.

$$E(Y_i|x_i) = \pi_i = \frac{exp(X_i'\beta)}{1 + exp(X_i'\beta)} \tag{13}$$

Em que, $X_i'$ representa a transposta da notação para $X_i$ e $\beta$ agora é um vetor com os coeficientes.

A transformação logito tem a seguinte forma e disponibiliza o seguinte resultado linearizado:

$$g(x) = \ln\left[\frac{\pi_i}{1-\pi_i}\right] = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1}x_{ip-1} \tag{14}$$

A estimação dos $p$ parâmetros também é feita com a mesma técnica da máxima verossimilhança, a qual também necessitará de cálculos numéricos para finalmente estimar valores para os $\beta_i$, tal que, $0 \leq i \leq p-1$, a resposta após igualar a zero as derivadas para cada $\beta$ será um vetor $px1$ com os parâmetros estimados $\hat{\beta}_i$. O modelo multiplo dependerá de mais de uma variável, mas para sua interpretação será levado em consideração apenas uma como é no caso simples, em que as outras serão mantidas constantes. Assim o modelo inclui os efeitos conjuntamente e também consegue estimar apenas com relação a uma variável explicativa.

Neste caso, também será encontrada uma matriz de covariâncias, que é obtida pela função de verossimilhança quando se maximiza duas vezes em $\beta_j$ e $\beta_l, j \neq l$. Será obtida uma matriz que é denotada por $I(\beta)$ e intitulada como matriz de observação e informação. Ao se calcular sua inversa é possível encontrar as variâncias e covariâncias dos estimadores, tal que na diagonal principal estão as variâncias e nas demais as covariâncias entre os parâmetros do modelo. Essa seria a formulação para essa matriz de estimações:

$$\hat{I}(\hat{\beta}) = X'VX \tag{15}$$

em que,

$$V = \begin{bmatrix} \hat{\pi_1}(1-\hat{\pi_1}) & 0 & \dots & 0 \\ 0 & \hat{\pi_2}(1-\hat{\pi_2}) & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & \hat{\pi_n}(1-\hat{\pi_n}) \end{bmatrix}$$

A partir deste resultado é possível calcular os intervalos com $(1-\alpha)\%$ de confiança, observando a raiz das variâncias estimadas. A formulação então será a mesma do caso simples.

### 2.1.2 Estimação da razão de chances

Em maior parte o objetivo deste tipo de estudo é definir uma probabilidade do desfecho ser acometido dado uma ou mais características. Assim o proposto é estimar a razão de prevalência. A regressão logistica não é capaz de fazer essa estimação mas consegue estimar uma medida que se assemelha: a razão de chances (OR) "*odds ratio*". Essa tem o mesmo intuito, porém dependendo de algumas situações ela se torna superestimada.

A fim de se obter uma fácil interpretação do modelo será agora considerada uma regressão logística simples a qual tem uma variável dicotômica como variável independente (explicativa). A OR estima essa prevalência da seguinte forma:

$$\hat{OR} = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} = \frac{\frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{\frac{1}{1+e^{\beta_0+\beta_1}}}}{\frac{\frac{e^{\beta_0}}{1+e^{\beta_0}}}{\frac{1}{1+e^{\beta_0}}}} = e^{\beta_1} \tag{16}$$

Observa-se na equação (16), que quando se obtêm a função da regressão logística para cada termo, após as simplificações tem-se que a $\hat{OR}$ é exatamente a exponencial do parâmetro de interesse. É possível assim calcular a chance de um indivíduo ter o desfecho presente dado uma característica x que está relacionada a esse parâmetro estimado. Logo, para a estimativa direta da OR apenas é necessário aplicar o logaritmo nos parâmetros. A OR é totalmente assimétrica à direita pois está entre o intervalo de 0 e $+\infty$, além disso quando se têm como resposta o valor 1 significa que não há diferença entre as classificações da variável resposta com relação a ter ou não o desfecho, portanto este é o valor nulo para a estimação. Mesmo com essa assimetria, para amostras suficientemente grandes ela tende a uma distribuição normal. Porém, na maioria dos casos, principalmente por se tratar de estudos epdemiológicos, os quais possuem muitos limites para seus desenvolvimentos, como por exemplo, amostras pequenas.

Como detalhado anteriormente, é fácil, por meio de softwares, definir os intervalos de confiança para os parâmetros e assim, consequentemente, também é fácil obter o intervalo de $(1-\alpha)\%$ de confiança para a $\hat{OR}$. É necessário apenas fazer o cálculo da exponêncial do intervalo já calculado para $\beta_i$ conforme a equação.

$$exp[\hat{\beta}_i \pm z_{1-\frac{\alpha}{2}} \hat{se}(\hat{\beta}_i)] \tag{17}$$

Se utilizará também o quantil da distribuição normal de acordo com a significância necessária e pode-se afirmar que a maior parte dos intervalos excederão o valor um, já que este é considerado o valor nulo para a OR.

### 2.1.3 Tipos de variáveis explicativas

Como a melhor explicação do modelo depende de novas variáveis que são agregadas a ele, é importante uma abordagem sobre como serão incluídas e interpretadas todos os tipos de variáveis explicativas, como as qualitativas, com mais de duas classificações, ordinal ou nominal e as quantitativas contínuas e discretas. Qualquer dessas podem ser incluídas e são capazes de ressaltar ainda mais as conclusões obtidas para a razão de prevalência. As variáveis qualitativas não terão diferença com relação a sua classificação, nominal e ordinal, mas, com relação à quantidade de classes, existirá um certo aspecto que mudará na construção do modelo.

Até o momento foi apresentado apenas um tipo de variável explicativa, que é classificada como dicotômica, ou seja, possui apenas duas classes. Para a inserção de uma variável explicativa politômica, a qual possui $k$ níveis, com $k \geq 3$, é necessária a criação de variáveis *designers*, ou mais comumente chamadas de *"dummy variables"* em que são criadas variáveis indicadoras $D_{lj}$ para os $k_j - 1$ níveis das variáveis originais. Escolhe-se um nível de referência, podendo ter um porque da escolha ou não, e então se codificam essas novas variáveis indicadoras com valores iguais a 0, para os outros níveis, ou seja, para as outras variáveis indicadoras poderão receber o valor 1 quando o indivíduo não estiver no nível de referência e pertencer a classe que esta variável representa. A linearização do novo modelo pode ser descrito da seguinte forma:

$$g(x) = \beta_0 + \beta_1 x_1 + \ldots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \beta_p x_p \tag{18}$$

Em que, existe uma variável indicadora $D$ e um coeficiente $\beta$ para os $k_j - 1$ níveis da j-ésima variável politõmica com $l$ níveis. Portanto, a diferença é a existência de mais variáveis explicativas no modelo, uma para cada nível da variável explicativa.

As estimações dos parâmetros são definidas da forma usual, por máxima verossimilhança. Além disso as estimações pontuais e intervalares para a razão de chances também são calculadas da mesma forma. Portanto continua a valer a relação entre as estimações de OR e os coeficientes $\beta$,ou seja, $ln(\hat{OR}) = \beta_i$. Agora serão comparadas dois a dois níveis da variável que se tornou indicadora segundo o desfecho, calculando a chance do indivíduo por exemplo ter o desfecho, ou seja, a doença dado um dos níveis da variável politômica, esta pode ser mensurada e confrontada com outro nível pré-estabelecido desta variável.

Outra classificação para variáveis explicativas são as quantitativas. Tanto as discretas como as contínuas podem ser tratadas de formas similares. Para um conjunto grande de diferentes valores, na maioria dos casos não se dá prioridade para uma categorização deste tipo de variável. Portanto, para uma interpretação mais efetiva deve-se determinar uma constante $c$ de incremento ou decremento, a qual dará um maior sentido na análise. As estimações também ocorrerão da mesma forma e pode-se obter tal relação:

$$ln(O\hat{R}(c)) = c\beta_i \tag{19}$$

Assim, a $\hat{OR}$ será interpretada como a chance de um indivíduo ter a doença com uma quantidade c da variável explicatíva contínua comparado com todos os outros possíveis valores de c. Outra fator importante é a equação (22), que representa a estimação intervalar da OR para este tipo de variável.

$$exp[c\hat{\beta}_i \pm z_{1-\frac{\alpha}{2}} c\hat{se}(\hat{\beta}_i)] \tag{20}$$

Observa-se que o princípio do cálculo é o mesmo. É necessário apenas definir a constante e inclui-la nas estimações. Assim, para todos os tipos de variáveis explicativas existe a mesma incoação para as estimações dos parâmetros a fim de aferir os valores para a $\hat{OR}$ que é o interesse maior neste estudo.

## 2.2 Regressão de Poisson

Essa metodologia está cada vez mais usual, principalmente em estudos epdemiológicos, em que existe uma melhor consistência das estimações e uma convergência certa do modelo. A regressão de Poisson é útil quando a variável resposta é uma contagem, por exemplo, uma contagem de doentes.

O modelo deve levar em consideração uma proporção, ou seja, uma medida relativa de acordo com o número de ocorrências especificado para um determinado intervalo, tal que t pode representar essa relação,

$$t = \frac{\text{número de eventos de interesse}}{\text{unidades totais}} \tag{21}$$

Geralmente as unidades totais, ou a referência, podem ser relacionadas ao espaço, tempo ou quantidade total dos possíveis casos. Essa variável da equação (21) também pode ser chamada de taxa de incidência e deve ser inserida, pois unidades amostrais podem ser acompanhadas em diferentes tempos e espaços.

Assim, Y terá distribuição Poisson, por se tratar de uma contagem e possui a seguinte distribuição de probabilidade:

$$f(Y) = \frac{(t\mu)^Y exp(-t\mu)}{Y!} \tag{22}$$

Como este estudo é de cunho transversal, pode-se atribuir o valor unitário ao tempo associado a cada participante (Coutinho et. al, 2008). A função que representa a regressão de Poisson é definida da seguinte forma:

$$Y_i = E(Y_i) + \varepsilon_i \quad i = 1, 2, 3, .., n \tag{23}$$

Com $E(Y_i) = e^{x_i'\beta}$.

A esperança para cada i caso, será denotada por $\mu_i$ e para relacioná-la com as variáveis explicativas $X_1, ..., X_{p-1}$ tem-se a seguinte notação $\mu(X_i, \beta)$. Como na regressão logística, existe também uma função de ligação para a linearização do modelo. Além disso, é garantido uma média de casos não negativa.

$$log(E(Y_i)) = log(\mu_i) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \tag{24}$$

Os parâmetros do modelo são estimados pelo método da quase verossimilhança, equação (25). Essa estima por sua vez o logaritimo da razão de prevalências (RP), dado por $log(\hat{\beta})$. A equação está para uma estimação de um modelo simples para uma melhor compreesão.

$$l(\beta_0, \beta_1) = C \sum_{i=1}^{n} [y_i(\beta_0 + \beta_1 x_i) - exp(\beta_0 + \beta_1 x_i)] \tag{25}$$

De acordo com Zou, (2003) pode ocorrer um erro para a estimação do intervalo de confiança dos parâmetros do modelo, em que, por suposição, o erro do modelo tem distribuição de Poisson, o que na verdade, o mais correto, seria ter uma distribuição binomial. O autor propõe uma estimação de variância robusta, a qual corrigirá a precisãao do intervalo de confiança. O estimador é chamado de *sandwich* e então serão consideradas as devidas correções. Algumas programações, como é o caso do software SAS, já contam com uma estimação robusta para a variância dos parâmetros e é por esse motivo que a estimação é feita pelo método de quase verossimilhança e estima uma aproximação para a RP.

A razão de prevalência é calculada da seguinte forma quando se têm um modelo simples, o que se torna possível observar que a mesma é estimada pelo logarítmo dos parâmetros estimados de acordo com a variável independente que se deseja estudar.

$$RP = \frac{\pi(1)}{\pi(0)} = e^{(1-0)\beta_1}$$

Dada a estimação dos parâmetros e de suas variâncias é possível encontrar o intervalo de confiança para o mesmo, já que o interesse é uma estimação da razão de prevalência (RP) obtida pela exponecial dessas estimações.

$$IC(\hat{RP}) = exp[\hat{\beta}_i \pm z_{1-\frac{\alpha}{2}} \hat{se}(\hat{\beta}_i)] \tag{26}$$

Por fim, todas as outras aplicações com relação aos tipos de variáveis explicativas vistas na regressão logística também são válidas na regressão de Poisson.

## 2.3 Regressão Log-binomial

Pelo modelo de regressão Log-binomial, é possível estimar diretamente o coeficiente que traduz a associação entre a variável resposta dicotômica e suas respectivas variáveis explicativas, ou seja, aquelas consideradas como de confusão. Esta associação é explicitada pelo razão de prevalência (RP). Neste, pode-se inferir uma prevalência a favor, contra ou igual entre dois grupos de interesse.

O modelo segue a mesma construção da regressão logística, partindo do princípio de erros com distribuição binomial como base para os cálculos. A diferença está na função de ligação, que é determinada por uma função logaritmica. O modelo é definido pela seguinte equação:

$$P(Y=1|X=x_i) = \pi(x_i) = e^{x_i'\beta} \tag{27}$$

E a função de ligação é dada por,

$$ln(\pi(x_i)) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \tag{28}$$

A fim de limitar as estimações para as proporções permanecerem entre 0 e 1, é necessário que $\sum_{l=1}^{p-1} \beta_p < 0$. Assim o espaço de parâmetros se torna restrito para a completa explicação da variável resposta.

A estimação dos parâmetros é feita pelo método da máxima verossimilhança, garantindo a escolha do melhor estimador com menor variância possível. As estimações também necessitam de métodos interativos para suas definições e em algumas vezes existem dificuldades de convergência, ou seja, ao usar o método de estimação de máxima verossimilhança juntamente com métodos interativos, não se encontra um estimativa para o parâmetro.

Williamson et. al (2013), cita três ocasiões em que não há convergência da estimação, são elas: na fronteira do espaço dos parâmetros (isto é, onde o preditor linear é igual a 0); no limite (isto é, quando o preditor linear se dirige para $-\infty$); ou dentro do espaço dos parâmetros.

O autor também explica que, na maioria dos casos de não convergência, uma solução prática é a reparametrização do modelo ajustado de acordo com uma das regiões citadas e que, além disso, a problemática encontrada não está no modelo de fato e sim nos pacotes dos programas que não são totalmente confiáveis. Além disso, alguns parâmetros desses pacotes também podem ser mudados, ou seja, aumentar o número de interações, ou alterar os valores iniciais dos preditores lineares.

A função de maxima verossimilhança deverá ser maximizada e igualada a zero. A equação (29) define a mesma para o j-ésimo beta estimado:

$$\sum_{i=1}^{n} x_{ij} \left( \frac{y_i - \pi(x_j)}{1 - \pi(x_j)} \right) = 0 \tag{29}$$

Como a função de ligação é a mesma do modelo de Poisson, as estimativas pontuais e intervalares da RP serão calculadas da mesma forma descrita no modelo anterior.

## 2.4 Modelos com estrutura hierárquica entre as variáveis

A modelagem hierarquica consiste em agrupar determinadas variáveis que pertencem a um mesmo fim e que de certa forma possuem uma ordem que pode classificá-las. É interessante para a resolução de problemas que envolvem várias características que estão associadas, pode-se então dividir, em etapas, ou seja, em blocos de variáveis, e definir para cada bloco as variáveis que mais contribuem para explicar a variável de interesse. No presente estudo serão coletadas informações que suprirão três grupos, tal que cada estrutura está contida em outra com uma abrangência maior de forma encadeada.

Um estudo feito por Pazó et.al (2013) deseja encontrar melhores associações para o indicador ICSAP (Internações por Condições Sensíveis à Atenção Primária). O mesmo pode ser considerado como uma medida para avaliação da atenção primária à saúde e assim averiguar a qualidade de serviços ambulatoriais no estado do Espírito Santo.

A estrutura proposta possui três níveis: distal, intermediário e proximal. O autor agregou na primeira estrutura as variáveis geográficas e demográficas, situação econômica e estrutura social. Para o nível intermediário foram consideradas modelos de assistência à saúde e organização dos serviços de saúde. Por fim, no nível proximal, têm-se indicadores de desempenho do sistema de saúde. A análise foi feita separadamente para cada nível, para assim averiguar as variáveis de maior associação. Após, foram agrupados todos os níveis considerando as variáveis de maior explicação para o modelo.

Desta mesma forma este trabalho propôs incluir uma estrutura hierárquica das variáveis. Inicialmente, como nível distal, também estão contidas variáveis demográficas das mães pesquisadas. Para o nível intermediário têm-se algumas variáveis relacionadas a eventos traumáticos que ocorreram antes do parto e que o pesquisador acredita na existência de uma correlação. Como bloco para variáveis consideradas proximais serão avaliadas aqueles eventos que ocorreram durante o parto, ou seja, variáveis mais associadas com a depressão pós-parto. No primeiro bloco, todas as variáveis estão contidas no modelo. Após a avaliação das mais significativas agregou-se essas ao segundo bloco e assim também foi feito para o terceiro bloco.

O uso desse tipo de estrutura facilita para uma melhor organização dos fatores que podem contribuir para o modelo de tal forma que seja possível inserir vários níveis de abrangência dentro de um modelo e assim melhorar sua capacidade de explicação da variável resposta.

# 3 Banco de dados

O banco de dados refere-se a mulheres grávidas internadas em uma UTI neonatal. O interesse do estudo é analisar a prevalência de depressão pós-parto nessas mães. Necessitou-se de alguns atributos que podem contribuir para explicar melhor essa variável. Como dito, foram coletadas informações referentes a situações socio-demográficas, antes do parto e situações após o nascimento e durante a internação de seus filhos. Acredita-se que essas três classes de variáveis explicativas estejam correlacionadas com a variável de interesse e encadeadas entre si e formando a estrutura tratada na seção anterior.

Pelo fato de considerar apenas as variáveis de interesse, o banco de dados utilizado para essa análise contém 372 mães, que foram incluídas na pesquisa nos três blocos de informação. Será feita uma análise inicial para se constatar as prevalências de depressão pós-parto para cada variável explicativa de interesse e em seguida as três modelagens propostas, focando na estimação de seus parâmetros em que resultará na medida de interesse do estudo.

## 3.1 Análise dos dados

Serão determinadas as variáveis que estarão contidas nos blocos descritos, em seguida serão feitas análises com relação a prevalência da doença para cada variável explicativa de cada bloco em que estão organizadas da seguinte forma:

- Bloco 1 - socio demográficas (distal):
  - Escolaridade da mãe.
  - Renda familiar.
  - Idade da mãe.
  - Estado civil da mãe.
- Bloco 2 - eventos traumáticos antes do parto (intermediário):
  - Presença quando ocorreu um incêndio ou explosão.
  - Sofreu algum abuso ou violência de algum familiar.
  - Sofreu abuso físico quando criança.

- Viu alguém sofrer crime violento ou lesão grave.
- Sofreu alguma outra situação que ameaçou a vida.
- Alguma pessoa que ama morreu de forma violenta ou inesperada.
- Vivenciou outro evento violento.
- Aconteceu algo grave que você não consegue falar.
- Sofreu alguma experiência sexual não desejada ou violenta.
- Sofreu algum acidente grave.

- Bloco 3 - eventos traumáticos durante o parto (proximal):

  - Medo intenso de morrer e do bebê também.
  - Dor intensa e prolongada.
  - Achou a assistência da equipe de saúde inadequada.
  - Informada quanto aos procedimentos a ser submetida.
  - Sensação de perda de controle.
  - Sofreu alguma experiência humilhante.

Abaixo, encontra-se tabeladas as prevalências para a existência de DPP (depressão pós parto) segundo as variáveis explicativas, estão presentes estimações pontuais e intervalares.

Tabela 1: Frequência para as variáveis explicativas e estimações das respectivas prevalências

| **Variáveis** | **Frequência (n = 275)** | **Prevalência de DPP (%)** | **IC 95%** |
|---|---|---|---|
| **Escolaridade** | | | |
| Fundamental e Médio | 181 | 19,89 | 14,04; 25,74 |
| Superior | 94 | 20,21 | 12,04; 28,38 |
| **Renda (salários mínimos)** | | | |
| Até 1 | 75 | 22,67 | 13,13; 32,20 |
| Acima de 1 | 200 | 19,00 | 13,53; 24,27 |
| **Idade da mãe (anos)** | | | |
| 14 a 30 | 181 | 15,47 | 10,17; 20,77 |
| 31 a 44 | 94 | 28,72 | 19,52; 37,93 |
| **Estado civil** | | | |
| Solteira | 85 | 20,00 | 11,44; 28,56 |
| Não solteira | 190 | 20,00 | 14,28; 25,72 |
| **Incêndio ou explosão** | | | |
| Sim | 12 | 25,00 | 11,44; 28,56 |
| Não | 263 | 19,77 | 14,93; 24,61 |
| **Abuso ou violência** | | | |
| Sim | 17 | 41,18 | 17,63; 64,72 |
| Não | 258 | 18,60 | 13,83; 23,38 |
| **Abuso quando criança** | | | |
| Sim | 24 | 37,50 | 18,01; 56,99 |
| Não | 251 | 18,33 | 13,51; 23,14 |
| **Algum crime** | | | |
| Sim | 49 | 28,57 | 15,84; 41,30 |
| Não | 226 | 18,14 | 13,09; 23,20 |
| **Ameaça a vida** | | | |
| Sim | 34 | 29,41 | 14,00; 44,82 |
| Não | 241 | 18,67 | 13,72; 23,62 |

Tabela 2: Frequência para as variáveis explicativas e estimações das respectivas prevalências

| Variáveis | Frequência (n = 275) | Prevalência de DPP (%) | IC 95% |
|---|---|---|---|
| **Medo de morrer** | | | |
| Sim | 135 | 29,63 | 21,88; 37,38 |
| Não | 140 | 10,71 | 5,56; 15,87 |
| **Dor intensa** | | | |
| Sim | 176 | 21,59 | 15,47; 27,71 |
| Não | 99 | 17,17 | 9,70; 24,65 |
| **Equipe inadequada** | | | |
| Sim | 29 | 17,24 | 3,41; 31,08 |
| Não | 246 | 20,32 | 15,26; 25,38 |
| **Informação procedente** | | | |
| Sim | 240 | 18,33 | 13,41; 23,26 |
| Não | 35 | 31,43 | 15,95; 46,90 |
| **Perda de controle** | | | |
| Sim | 77 | 29,87 | 19,58; 40,16 |
| Não | 198 | 16,16 | 11,00; 21,32 |
| **Experiência humilhante** | | | |
| Sim | 8 | 62,50 | 28,74; 96,26 |
| Não | 267 | 18,73 | 14,01; 23,43 |
| **Morte inesperada** | | | |
| Sim | 101 | 27,72 | 18,94; 36,51 |
| Não | 174 | 15,52 | 10,10; 20,93 |
| **Evento violênto** | | | |
| Sim | 23 | 21,74 | 4,78; 38,70 |
| Não | 252 | 19,84 | 14,87; 24,80 |
| **Evento grave e não fala** | | | |
| Sim | 18 | 38,89 | 16,23; 61,55 |
| Não | 257 | 18,68 | 13,88; 23,47 |
| **Experiência sexual** | | | |
| Sim | 18 | 38,89 | 16,23; 61,55 |
| Não | 257 | 18,68 | 13,88; 23,47 |
| **Acidente grave** | | | |
| Sim | 8 | 25,00 | 0,00; 55,19 |
| Não | 267 | 11,61 | 7,74; 15,48 |

O estudo visa compreender os limites encontrados pela estimação dos parâmetros realizada pela regressão logística, o que está diretamente relacionado com a prevalência entre as categorias das variáveis explicativas. A partir dos quadros 1 e 2, destacam-se diferenças de prevalências que consequentemente acarretarão estimações viesadas da razão de chances.

Para o bloco 1 observa-se que a prevalência de DPP em mães com idade de 31 a 44 anos é aproximadamente o dobro da prevalência das mães com 14 a 30 anos, 28,72 e 15,47, respectivamente. Para as outras variáveis não se observou tanta discrepância nos valores de prevalência de DPP. No bloco 2 existe uma diferença notável entre as prevalências da variável abuso ou violência, com uma prevalência de DPP de 41,18 para as mães que sofreram abuso e 18,60 para aquelas que não sofreram abuso. Outra variável com diferença perceptível é a morte inesperada de pessoa próxima com uma prevalência de DPP de 27,72 para quem passou por esse evento e um a prevalência de 15,52 para as mães que não sofreram esse evento.

Por fim o bloco 3 apresenta como variáveis de destaque medo de morrer, perda de controle e experiência humilhante, tal que a primeira possui uma prevalência de ter DPP quando as mães apresentaram medo de morrer sendo o triplo da prevalência de possuir DPP para as mães que não tinham medo de morrer ou que seu bebê morresse (29,63;10,71). A prevalência de ter a doença entre mães que relataram perda de controle foi de 29,87 contra uma prevalência de ter DPP em mães que não informaram esse evento de 16,16. A terceira variável, considerada como um evento ocorrido durante o parto, apresenta também o triplo da prevalência para as mães que passaram por esse evento com relação aquelas que não se aplicam nessa variável, com RP = 62,50 e RP = 18,73, respectivamente.

### 3.1.1 Análise com modelo de regressão logística

Nesta seção estão os resultados para a modelagem logística binária e múltipla, com a inserção do método hierárquico descrito. Serão apresentados os três blocos de variáveis com as estimações da OR para cada modelo. As variáveis foram selecionadas para prosseguirem dentro do modelo com uma significância de 0,1. Foram realizados testes de Wald e observado o p-valor das mesmas dentro do modelo múltiplo para assim tomar a decisão de permanência no modelo.

Para o bloco 1, tem-se o seguinte modelo múltiplo,

$$\pi(x) = \beta_0 + X_{i1}(Escolaridade) + X_{i2}(Renda) + X_{i3}(Idade) + X_{i4}(Estado \quad civil)$$

Tabela 3: Estimação da razão de chances (OR) para o modelo de regressão logística simples (bruta) e múltipla (ajustada)

Bloco 1 - Variáveis sócio-demográficas

| Variáveis | OR Bruta | | OR Ajustada | |
|---|---|---|---|---|
| | OR (IC 95%) | p-valor | OR (IC 95%) | p-valor |
| **Escolaridade** | | | | |
| Fundamental e médio | 1 | - | 1 | - |
| Superior | 1,02 (0,55; 1,90) | 0,9493 | 1,15 (0,59; 2,23) | 0,6789 |
| **Renda em salários mínimos** | | | | |
| Até 1 | 1 | - | 1 | - |
| Acima de 1 | 0,8 (0,42; 1,53) | 0,4989 | 0,73 (0,37; 1,45) | 0,3718 |
| **Idade da mãe (anos)** | | | | |
| 14 a 30 | 1 | - | 1 | - |
| 31 a 44 | 2,20 (1,21; 4,02) | 0,0101 | 2,27 (1,24; 4,15) | 0,0077 |
| **Estado civil** | | | | |
| Solteira | 1 | - | 1 | - |
| Não solteira | 1 (0,53; 1,89) | 1 | 0,91 (0,46; 1,77) | 0,7756 |

Para o bloco 2, agregou-se as variáveis desse nível, aquela do bloco 1 que apresentou ser significativa, assim o modelo é dado por,

$$\pi(x) = \beta_0 + X_{i1}(Idade) + variáveis \quad do \quad bloco \quad 2$$

Tabela 4: Estimação da razão de chances (OR) para o modelo de regressão logística simples (bruta) e múltipla (ajustada)

Bloco 2 - Eventos traumáticos anteriores ao parto

| Variáveis | OR Bruta | | OR Ajustada | |
|---|---|---|---|---|
| | OR (IC 95%) | p-valor | OR (IC 95%) | p-valor |
| **Incêndio / Explosão** | | | | |
| Sim | 1,35 (0,35; 5,17) | 0,6590 | 1,44 (0,49; 4,21) | 0,50 |
| Não | 1 | - | 1 | - |
| **Abuso / Violência** | | | | |
| Sim | 3,06 (1,11; 8,45) | 0,0308 | 1,53 (0,62; 3,74) | 0,3524 |
| Não | 1 | - | 1 | - |
| **Abuso quando criança** | | | | |
| Sim | 2,05 (1,15; 3,65) | 0,0153 | 1,55 (0,71; 3,39) | 0,2696 |
| Não | 1 | - | 1 | - |
| **Crime / Lesão grave** | | | | |
| Sim | 1,57 (0,93; 2,65) | 0,882 | 1,32 (0,72; 2,43) | 0,3724 |
| Não | 1 | - | 1 | - |
| **Ameaça a vida** | | | | |
| Sim | 1,57 (0,88;2,82) | 0,1270 | 1,06 (0,51; 2,21) | 0,8674 |
| Não | 1 | - | 1 | - |
| **Morte de pessoa próxima** | | | | |
| Sim | 1,79 (1,12; 2,85) | 0,0152 | 1,57 (0,96; 2,59) | 0,0732 |
| Não | 1 | - | 1 | - |
| **Outro evento violênto** | | | | |
| Sim | 1,09 (0,48; 2,47) | 0,8259 | 0,72 (0,34; 1,53) | 0,4154 |
| Não | 1 | - | 1 | - |
| **Algo grave e não consegue falar** | | | | |
| Sim | 2,08 (1,11; 3,92) | 0,0231 | 1,59 (0,80; 3,15) | 0,1822 |
| Não | 1 | - | 1 | - |
| **Experiência sexual ou violênta** | | | | |
| Sim | 2,08 (1,11; 3,92) | 0,0231 | 1,79 (0,93; 3,43) | 0,2325 |
| Não | 1 | - | 1 | - |
| **Acidente Grave** | | | | |
| Sim | 0,62 (0,10; 3,93) | 0,6100 | 0,35 (0,06; 1,97) | 0,2325 |
| Não | 1 | - | 1 | - |

No bloco 3, agregou-se as variáveis dessa classe e aquelas que apresentaram significância para o modelo, este é definido por,

$$\pi(x) = \beta_0 + X_{i1}(Idade) + x_{i2}(Morte \quad de \quad pessoa \quad próxima) + variáveis \quad do \quad bloco \quad 3$$

Tabela 5: Estimação da razão de chances (OR) para o modelo de regressão logística simples (bruta) e múltipla (ajustada)

Bloco 3 - Eventos traumáticos durante o parto

| Variáveis | OR Bruta | | OR Ajustada | |
|---|---|---|---|---|
| | OR (IC 95%) | p-valor | OR (IC 95%) | p-valor |
| **Medo de morrer** | | | | |
| Sim | 3,51 (1,83; 6,72) | 0,0002 | 2,78 (1,38; 5,62) | 0,0043 |
| Não | 1 | - | 1 | - |
| **Dor intensa** | | | | |
| Sim | 1,33 (0,70; 2,50) | 0,3801 | 1,26 (0,62; 2,56) | 0,5192 |
| Não | 1 | - | 1 | - |
| **Assistência inadequada** | | | | |
| Sim | 0,82 (0,30; 2,25) | 0,6950 | 0,49 (0,15; 1,63) | 0,2464 |
| Não | 1 | - | 1 | - |
| **Informação dos procedimentos** | | | | |
| Sim | 0,50 (0,22; 1,07) | 0,0747 | 0,48 (0,18; 1,27) | 0,1379 |
| Não | 1 | - | 1 | - |
| **Perda de controle** | | | | |
| Sim | 2,21 (1,19;4,10) | 0,0119 | 1,43 (0,68; 3,02) | 0,3459 |
| Não | 1 | - | 1 | - |
| **Experiência humilhante** | | | | |
| Sim | 7,23 (1,67; 31,27) | 0,0081 | 4,91 (1,34; 17,93) | 0,0160 |
| Não | 1 | - | 1 | - |

Após a análise dos quadros 3, 4 e 5 pode-se inferir informações sobre os três blocos hierárquicos observados usando uma modelagem de regressão logística. Para as variáveis sócio demográficas, observa-se que a variável idade da mãe possui uma maior

OR estimada dentre as variáveis no modelo múltiplo, ou seja, a chance de uma mãe com 31 a 44 anos ter DPP é 2,27 vezes a chance de uma mãe possuir a doença quando tem 14 a 30 anos. Com um intervalo de 95% de confiança de 1,24 a 4,15 há chance de ter DPP quando uma mãe possui 31 a 44 anos. Observa-se também que todas as outras variáveis são independentes de ter ou não a doença, ou seja, para qualquer classe das variáveis escolaridade, renda e estado civil não existe diferença para mães com a presença de DPP, já que o intervalo de confiança contém o número 1. Além disso, a única variável significativa para o modelo é a idade, portanto para os próximos blocos, essa variável já estará incluída no modelo múltiplo.

Dentre as variáveis de eventos traumáticos antes do parto, apenas uma se revelou interessante para a agregação no bloco seguinte, morte de pessoa próxima, com um p-valor muito próximo da fronteira de 0,1 (0,0732). Seu intervalo de confiança, assume o valor 1, portanto essa variável independe da existência de DPP, assim como todas as variáveis desse bloco.

Dessa mesma forma, o grupo de variáveis do bloco 3 também foi analisado, concluindo que duas das variáveis são significativas: medo de morrer ou que seu bebê morra e experiência humilhante. Além disso, infere-se que a chance de uma mãe ter DPP e ter sentido medo de morrer ou a morte de seu bebê é 2,78 a chance de ter DPP em mães que não obtiveram esse sentimento e a chance de ter DPP em uma mãe que passou uma experiência humilhante é de 4,91 vezes a chance de uma mãe que não passou por uma experiência desse tipo e apresentar DPP.

Em geral, as OR estimadas para o modelo simples e múltiplo não sofrem tanta alteração, salvo alguns casos, mas sua significância e intervalos de confiança possuem alteração quando são analisadas conjuntamente com as outras variáveis, além de possuir intervalos de grande amplitude.

### 3.1.2 Análise com o modelo de regressão de Poisson

Conforme feito na seção anterior, serão modeladas as mesmas variáveis, levando em consideração a estrutura hierárquica proposta. Sendo definidas agora a estimação para a razão de prevalência (RP) com variância robusta.

Tabela 6: Estimação da razão de prevalência (RP) para o modelo de regressão Poisson simples (bruta) e múltipla (ajustada)

Bloco 1 - Variáveis sócio-demográficas

| Variáveis | RP Bruta | | RP Ajustada | |
|---|---|---|---|---|
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Escolaridade** | | | | |
| Fundamental e médio | 1 | - | 1 | - |
| Superior | 1,02 (0,62; 1,67) | 0,9493 | 1,12 ( 0,66; 1,88) | 0,6759 |
| **Renda ( em SM)** | | | | |
| Até 1 | 1 | - | 1 | - |
| Acima de 1 | 0,84 (0,5; 1,39) | 0,4948 | 0,78 (0,46; 1,33) | 0,3672 |
| **Idade da mãe (anos)** | | | | |
| 14 a 30 | 1 | - | 1 | - |
| 31 a 44 | 1,86 (1,16; 2,96) | 0,0093 | 1,90 (1,19; 3,02) | 0,0069 |
| **Estado civil** | | | | |
| Solteira | 1 | - | 1 | - |
| Não solteira | 1,00 (0,60; 1,67) | 1,0000 | 0,93 (0,55; 1,56) | 0,7760 |

No bloco 2 (tabela 7), como aconteceu para regressão logística, agregou-se as variáveis significativas no modelo múltiplo e consequentemente obteve-se o mesmo modelo.

$$\pi(x) = \beta_0 + X_{i1}(Idade) + variáveis \quad do \quad bloco \quad 2$$

Tabela 7: Estimação da razão de prevalência (RP) para o modelo de regressão de Poisson simples (bruta) e múltipla (ajustada)

Bloco 2 - Eventos traumáticos anteriores ao parto

| Variáveis | RP Bruta | | RP Ajustada | |
|---|---|---|---|---|
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Incêndio / Explosão** | | | | |
| Sim | 1,26 (0,46; 3,47) | 0,6488 | 1,43 (0,48; 4,27) | 0,5213 |
| Não | 1 | - | 1 | - |
| **Abuso / Violência** | | | | |
| Sim | 2,21 (1,19; 4,13) | 0,0124 | 1,13 (0,43; 2,99) | 0,8010 |
| Não | 1 | - | 1 | - |
| **Abuso quando criança** | | | | |
| Sim | 2,05 (1,15; 3,65) | 0,0153 | 1,47 (0,66; 3,28) | 0,3443 |
| Não | 1 | - | 1 | - |
| **Crime / Lesão grave** | | | | |
| Sim | 1,57 (0,93; 2,65) | 0,0882 | 1,44 (0,81; 2,55) | 0,2159 |
| Não | 1 | - | 1 | - |
| **Ameaça a vida** | | | | |
| Sim | 1,57 (0,89; 2,82) | 0,1270 | 0,90 (0,44; 1,85) | 0,7739 |
| Não | 1 | - | 1 | - |
| **Morte de pessoa próxima** | | | | |
| Sim | 1,79 (1,12; 2,85) | 0,0152 | 1,68 (1,04; 2,73) | 0,0348 |
| Não | 1 | - | 1 | - |
| **Outro evento violênto** | | | | |
| Sim | 1,10 (0,48; 2,47) | 0,8259 | 0,66 (0,31; 1,38) | 0,4104 |
| Não | 1 | - | 1 | - |
| **Algo grave e não consegue falar** | | | | |
| Sim | 2,08 (1,11; 3,92) | 0,0231 | 1,54 (0,75; 3,14) | 0,2345 |
| Não | 1 | - | 1 | - |
| **Experiência sexual ou violênta** | | | | |
| Sim | 2,08 (1,10; 3,92) | 0,0231 | 1,59 (0,80; 3,15) | 0,1824 |
| Não | 1 | - | 1 | - |
| **Acidente grave** | | | | |
| Sim | 0,62 (0,10; 3,93) | 0,6100 | 0,32 (0,06; 1,58) | 0,1616 |
| Não | 1 | - | 1 | - |

Por fim no quadro 8, agregou-se às variáveis do bloco 3 aquelas que apresentaram significância para o modelo, este é definido por,

$$\pi(x) = \beta_0 + X_{i1}(Idade) + X_{i2}(Morte \quad de \quad pessoa \quad próxima) + variáveis \quad do \quad bloco \quad 3$$

Tabela 8: Estimação da razão de prevalência (RP) para o modelo de regressão de Poisson simples (bruta) e múltipla (ajustada)

| Bloco 3 - Eventos traumáticos durante o parto | | | | |
|---|---|---|---|---|
| Variáveis | RP Bruta | | RP Ajustada | |
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Medo de morrer** | | | | |
| Sim | 2,76 (1,60; 4,77) | 0,0002 | 2,24 (1,25; 4,01) | 0,0069 |
| Não | 1 | - | 1 | - |
| **Dor intensa** | | | | |
| Sim | 1,26 (0,75; 2,11) | 0,3845 | 1,17 (0,69; 1,99) | 0,5699 |
| Não | 1 | - | 1 | - |
| **Assistência inadequada** | | | | |
| Sim | 0,85 (0,37; 1,95) | 0,6993 | 0,63 (0,28; 1,38) | 0,2472 |
| Não | 1 | - | 1 | - |
| **Informação dos procedimentos** | | | | |
| Sim | 0,58 (0,31; 1,02) | 0,0581 | 0,66 (0,35; 1,02) | 0,2472 |
| Não | 1 | - | 1 | - |
| **Perda de controle** | | | | |
| Sim | 1,85 (1,16; 2,95) | 0,0099 | 1,27 (0,74; 2,17) | 0,3883 |
| Não | 1 | - | 1 | - |
| **Experiência humilhante** | | | | |
| Sim | 3,34 (1,85; 6,03) | $\leq 0,0001$ | 1,83 (1,03; 3,25) | 0,0403 |
| Não | 1 | - | 1 | - |

A composição dos modelos, com relação a significância das variáveis é a mesma do modelo de regressão logística, tal que para cada bloco se inseriu as mesmas variáveis para o modelo múltiplo.

Como dito, a idade da mãe é agregada ao bloco seguinte e mostra que a chance de uma mãe ter DPP com 31 a 44 anos é 90% maior do que chance de uma mãe ter DPP com idade entre 14 e 30 anos. Dentre as variáveis anteriores ao parto, observa-se que a chance de ter DPP quando uma mãe sofreu alguma morte de pessoa próxima é 68% maior do que aquelas mães que não passaram por esse evento. No último bloco, tem-se que a chance de uma mãe possuir DDP quando a mesma possuiu medo de morrer ou de seu bebê morrer é de 2,24 vezes a chance das mães que não obtiveram este evento possuir DPP e a chance de uma mãe possuir a doença quando já passou por uma experiência humilhante é de 83% maior do que a chance das mães que não passaram por esse evento possuir a doença. Ao observar todas a estimações, entre o modelo simples e múltiplo, existem variáveis que deixaram de ser significativas dentro de um modelo com mais variáveis assim como aquelas que eram consideradas significativas no modelo simples também deixaram de ser no modelo múltiplo.

### 3.1.3 Análise com o modelo de regressão log-binomial

Será considerada a modelagem log-binomial para as mesmas estimações realizadas nas duas seções anteriores, essa regressão, conforme dito, é capaz de estimar diretamente a razão de prevalência (RP).

Tabela 9: Estimação da razão de prevalência (RP) para o modelo de regressão log-binomial simples (bruta) e múltipla (ajustada)

Bloco 1 - Variáveis sócio-demográficas

| Variáveis | RP Bruta | | RP Ajustada | |
|---|---|---|---|---|
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Escolaridade** | | | | |
| Fundamental e médio | 1 | - | 1 | - |
| Superior | 1,09 (0,69; 1,72) | 0,714 | 1,07 ( 0,65; 1,77) | 0,7815 |
| **Renda ( em SM) pendente** | | | | |
| Até 1 | 1 | - | 1 | - |
| Acima de 1 | 0,76 (0,48; 1,20) | 0,2385 | 0,76 (0,46; 1,26) | 0,2905 |
| **Idade da mãe (anos)** | | | | |
| 14 a 30 | 1 | - | 1 | - |
| 31 a 44 | 1,71 (1,10; 2,66) | 0,0115 | 1,78 (1,14; 2,78) | 0,0117 |
| **Estado civil** | | | | |
| Solteira | 1 | - | 1 | - |
| Não solteira | 1,03 (0,63; 1,68) | 0,916 | 1,00 (1,14; 2,78) | 0,9956 |

No bloco 2 ( *quadro 10*), agregou-se as variáveis e obteve-se também o mesmo modelo múltiplo.

$$\pi(x) = \beta_0 + X_{i1}(Idade) + variáveis \quad do \quad bloco \quad 2$$

Tabela 10: Estimação da razão de prevalência (RP) para o modelo de regressão log-binomial simples (bruta) e múltipla (ajustada)

Bloco 2 - Eventos traumáticos antes do parto

| Variáveis | RP Bruta | | RP Ajustada | |
|---|---|---|---|---|
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Incêndio / Explosão** | | | | |
| Sim | 1,01 (0,36; 2,83) | 0,981 | 1,09 (0,30; 2,76) | 0,8792 |
| Não | 1 | - | 1 | - |
| **Abuso / Violência** | | | | |
| Sim | 2,28 (1,33; 3,92) | 0,0027 | 1,55 (0,54; 2,46) | 0,7083 |
| Não | 1 | - | 1 | - |
| **Abuso quando criança** | | | | |
| Sim | 1,94 (1,13; 3,35) | 0,0163 | 1,51 (0,77; 2,99) | 0,2319 |
| Não | 1 | - | 1 | - |
| **Crime / Lesão grave** | | | | |
| Sim | 1,60 (0,99; 2,60) | 0,055 | 1,32 (0,83; 2,10) | 0,2319 |
| Não | 1 | - | 1 | - |
| **Ameaça a vida** | | | | |
| Sim | 1,60 (0,94; 2,72) | 0,0804 | 1,1 (0,59; 2,04) | 0,7603 |
| Não | 1 | - | 1 | - |
| **Morte de pessoa próxima** | | | | |
| Sim | 1,76 (1,14; 2,73) | 0,0111 | 1,74 (1,12; 2,75) | 0,0157 |
| Não | 1 | - | 1 | - |
| **Outro evento violênto** | | | | |
| Sim | 0,98 (0,43; 2,21) | 0,9600 | 0,70 (0,32; 1,49) | 0,3534 |
| Não | 1 | - | 1 | - |
| **Algo grave e não consegue falar** | | | | |
| Sim | 1,93 (1,03; 3,61) | 0,0393 | 1,59 (0,84; 3,01) | 0,1513 |
| Não | 1 | - | 1 | - |
| **Experiência sexual ou violenta** | | | | |
| Sim | 1,93 (1,03; 3,61) | 0,0393 | 1,27 (0,73; 2,23) | 0,3972 |
| Não | 1 | - | 1 | - |
| **Acidente grave** | | | | |
| Sim | 0,58 (0,09; 3,68) | 0,563 | 0,45 (0,08; 2,51) | 0,3654 |
| Não | 1 | - | 1 | - |

Agrupou-se às variáveis do bloco 3 (tabela 11) aquelas que também apresentaram significância para o modelo, este é definido por,

$$\pi(x) = \beta_0 + X_{i1}(Idade) + X_{i2}(Morte \quad de \quad pessoa \quad próxima) + variáveis \quad do \quad bloco \quad 3$$

Tabela 11: Estimação da razão de prevalência (RP) para o modelo de regressão log binomial simples (bruta) e múltipla (ajustada)

Bloco 3 - Eventos traumáticos durante o parto

| Variáveis | RP Bruta | | RP Ajustada | |
|---|---|---|---|---|
| | RP (IC 95%) | p-valor | RP (IC 95%) | p-valor |
| **Medo de morrer** | | | | |
| Sim | 2,64 (1,59; 4,40) | 0,0002 | 2,05 (1,21; 3,46) | 0,0075 |
| Não | 1 | - | 1 | - |
| **Dor intensa** | | | | |
| Sim | 1,38 (0,83; 2,28) | 0,214 | 1,20 (0,72; 2,01) | 0,4797 |
| Não | 1 | - | 1 | - |
| **Assistência inadequada** | | | | |
| Sim | 0,76 (0,33; 1,74) | 0,528 | 0,75 (0,39; 1,47) | 0,4051 |
| Não | 1 | - | 1 | - |
| **Informação dos procedimentos** | | | | |
| Sim | 0,61 (0,35; 0,97) | 0,0739 | 0,97 (0,72; 1,30) | 0,8345 |
| Não | 1 | - | 1 | - |
| **Perda de controle** | | | | |
| Sim | 1,99 (1,28; 3,09) | 0,0021 | 1,31 (0,80; 2,13) | 0,2777 |
| Não | 1 | - | 1 | - |
| **Experiência humilhante** | | | | |
| Sim | 3,14 (1,75; 5,66) | 0,0001 | 1,79 (1,11; 2,9) | 0,0171 |
| Não | 1 | - | 1 | - |

No modelo de regressão log-binomial algumas variáveis também sofreram alterações com relação às suas significâncias para um modelo simples e um modelo múltiplo, mas todas as variáveis interessantes para o modelo na análise múltipla também eram significativas no modelo simples. No primeiro bloco, como aconteceu nos modelos anteriores, a viriável idade da mãe foi inserida no modelo múltiplo do bloco 2, a mesma representa um RP de 1,78. Portanto a prevalência de DPP em mães de 31 a 44 anos é 78% maior do que em mães com 14 a 30 anos. No bloco 2 também se destaca a variável morte de pessoa próxima com um p-valor de 0,0157, tal que, a prevalência de se obter DPP em mães que passaram por esse evento é de 74% a mais do que em mães que não passaram por essa situação. Enfim, no bloco 3 encontram-se também as mesmas variáveis consideradas significativas nas outras modelagens, com uma prevalência de mais de 100% para a existência de DPP em mães que tem medo de morrer ou que seu bebê morra com relação às mães que não possuiram esse sentimento, para a variável experiência humilhante, obtêm-se uma prevalência de DPP de 79% nas mães que passaram por uma experiência humilhante após ou durante o parto com relação às mães que não passaram por essa experiência.

## 3.2 Comparação dos métodos

Com a criação de novas técnicas na análise de dados dicotômicos, observa-se que as mesmas conseguem lidar melhor com as limitações do modelo de regressão logística. Com relação à razão de prevalência (RP), o modelo de regressão logística estima a razão de chances (*odds ratio* OR) que pode substituir a RP, essa está em função da probabilidade do indivíduo ter ou não a doença e quando a prevalência do fator de risco se distância de 0, independente da categoria da variável que está ajudando a explicar, conclui-se a existencia de um viés que afeta as estimativas da razão de chances que consequentemente diverge da aproximação da razão de prevalência determinada por outras modelagens.

A OR pode ser representada por:

$$\hat{OR} = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \tag{30}$$

E pode ser reescrita da seguinte forma:

$$\hat{OR} = \frac{\pi(1)}{\pi(0)}\left(\frac{1-\pi(0)}{1-\pi(1)}\right) \cong RP\left(\frac{1-\pi(0)}{1-\pi(1)}\right) \tag{31}$$

Constata-se pela equação (31) que mesmo que a OR se assimile com a RP existe um viés que pode alterar e mascarar os valores estimados. Esse não acarretará problemas caso tenda ser igual a um, mas se alguma das probabilidades relacionadas às variáveis explicativas tenderem a valores muito diferentes de 0 o viés se distânciará de 1 e da estimação da RP assim a OR não poderá ser utilizada como referência para o estudo.

Na tabela 12, a seguir, estão as comparações entre os modelos com algumas variáveis que podem inserir um viés na estimação e definir essa disparidade entre a OR e a RP.

Tabela 12: Comparação das estimações de razão de chances e razão de prevalências segundo o modelo usado na forma múltipla e a estimação da prevalência obtida

| **Variáveis** | **Prevalência** | **Logística** | **Poisson** | **Log-binomial** |
|---|---|---|---|---|
| | | **OR (IC 95%)** | **RP (IC 95%)** | **RP (IC 95%)** |
| **Idade da mãe (anos)** | | | | |
| 14 a 30 | 15,47 | 1 | 1 | 1 |
| 31 a 44 | 28,72 | 2,27 (1,24; 4,15) | 1,90 (1,19; 3,02) | 1,78 (1,14; 2,7) |
| **Morte inesperada** | | | | |
| Não | 15,52 | 1 | 1 | 1 |
| Sim | 27,72 | 1,57 (0,96; 2,59) | 1,68 (1,04; 2,73) | 1,74 (1,12; 2,75) |
| **Abuso ou violência** | | | | |
| Não | 18,60 | 1 | 1 | 1 |
| Sim | 41,18 | 1,53 (0,62; 3,74) | 1,13 (0,43; 2,99) | 1,55 (0,54; 2,46) |
| **Medo de morrer** | | | | |
| Não | 10,71 | 1 | 1 | 1 |
| Sim | 29,63 | 2,78 (1,38; 5,62) | 2,24 (1,25; 4,01) | 2,05 (1,21; 3,46) |
| **Perda de controle** | | | | |
| Não | 16,16 | 1 | 1 | 1 |
| Sim | 29,87 | 1,43 (0,68; 3,02) | 1,27 (0,74; 2,17) | 1,31 (0,80; 2,13) |
| **Experiência humilhante** | | | | |
| Não | 18,73 | 1 | 1 | 1 |
| Sim | 62,50 | 4,91 (1,34; 17,93) | 1,83 (1,03; 3,25) | 1,79 (1,11; 2,9) |

A partir das prevalências observadas da existência de DPP, como as mesmas se distanciam de 0, observam-se valores diferentes para a estimação das ORs e RPs. Observa-se uma maior amplitude dos intervalos ao se usar a regressão logística, possivelmente as variâncias dos parâmetros foram superestimadas comparando com os intervalos dos outros modelos. Como existe uma prevalência alta para a variável experiência humilhante a mesma expressa uma maior divergência dos valores estimados pelos modelos, tal que a estimação pontual da OR é 62,7% maior do que a estimação da RP pelo modelo de Poisson, além de um intervalo muito discrepante de acordo com os outros modelos.

De acordo com Petersen e Deddens (2008) a regressão logística trabalha muito bem com relação às doenças raras, já que suas probabilidades serão baixas e não influenciarão no viés, quando as prevalências não ultrapassarem 10% a razão de chances é capaz de ser similar a razão de prevalência. Como a distribuição binomial pode ser aproximada para uma distribuição de Poisson com um tamanho de amostra grande e uma probabilidade pequena, assume-se a estimação da RP pelo modelo de Poisson. O autor definiu algumas simulações de prevalência para a comparação entre o modelo de Poisson e o log-binomial e concluiu que para prevalências baixas e intermediárias a modelagem log-binomial apresenta-se menos tendenciosa do que a regressão de Poisson, quando existem prevalências altas, acima de 50%, por exemplo, e uma maostra grande, acima de 1000, essas modelagens não sofrem muita disparidade, com uma diferença apenas na terceira ou quarta casa decimal da RP. Na tabela 12 a maior parte das prevalências estão em torno de 30, ou seja, as estimações log-binomial estarão menos tendenciosas, mas em maior parte se aproximam da estimação feita pelo modelo de Poisson, para as prevalências maiores também se encontram RPs parecidas entre os modelos, possívelmente não são iguais, como o autor coloca, pelo fato de uma amostra menor do que a que o mesmo faz referência.

Logo, pode-se inferir que para determinados casos a regressão logística superestima essa medida que é muito importante na epdemiologia e deve ser determinada com cautela pelo pesquisador, analisando sempre os dados descritivamente antes da escolha de qualquer modelagem. Além disso, têm-se algumas variações dentre os modelos alternativos, que também deverão ser definidos de acordo com as prevalências observadas caso o pesquisador deseje usar os mesmos.

# 4 Conclusão

O estudo tem como um dos objetivos determinar as modelagens para se estudar casos em que a resposta das variáveis são dicotômicas com uma abordagem teorica das possíveis relações com as variáveis explicativas, sendo elas, qualitativas e quantitativas. Foram apresentadas as particularidades das modelagens para este tipo de variável resposta, as quais são incluídas a um grupo maior, chamado de modelos generalizados. As modelagens vistas são denominadas, regressão logística, com um maior detalhamento, principalmente por conta da teoria já estar bem definida na literatura, regressão de Poisson e regressão log-binomial. Observou-se a relação entre os parâmetros e a medida principal resultante dessas regressões, que é a razão de prevalência. Infere-se que a mesma é capaz de ser estimada diretamente pela log-binomial, já a regressão logística estima a razão de chances que é similar. Porém, em alguns casos, superestima os resultados encontrados e a regressão de Poisson se aproxima melhor na maior parte das situações. Foram definidos aspectos relevantes principalmente correlacionados a essa medida, como as estimações dos parâmetros, pelo métodos da maxima varossimilhança, garantindo a eles variância mínima, distribuição normal e sendo o melhor e único estimador. Como consequência é possível gerar as inferências para uma população e o estudo se tornar válido para a mesma.

Diante disso, definiu-se estimações pontuais e intervalares para os parâmetros e suas devidas comparações para cada modelagem e cada conjunto de variáveis.

Usou-se um banco de dados real, relacionado à situações pisicologicas de mulheres que recentemente conceberam seus filhos. Foram avaliados vários aspectos para complementar o modelo e assim comparar as diversas situações com relação a essas três principais modelagens. Foi criado um ambiente que levou em consideração as três situações coletadas, variáveis relacionadas a situações traumáticas vividas, variáveis antes e duarnte o parto e variáveis socio-demográficas, as mesmas formando uma estrutura hierárquica em que estão encadeadas. Essa estrutura também foi levada em consideração na modelagem e na análise dos dados. Conclui-se que para a variável resposta, depressão pós-parto (desfecho), as variáveis que ajudam a explica-la são cada qual de um bloco determinado. Portanto, tem-se variáveis mais distantes, porém correlacionadas e variáveis mais diretamente correlacionadas, são elas: idade da mãe, com as categorias 14 a 30 anos e 31 anos a 44; como variável considerada em uma relação intermediária, tem-se a morte inesperada de uma pessoa próxima, com respostas sim e não, ou seja, se ocorreu ou não esse evento antes do parto e como variáveis mais próximas, foram selecionadas as variáveis medo de morrer ou que seu bebê morra e se essas mães passaram ou não por uma experiência humilhante no período que estavam no hospital com seus filhos recém-nascidos. Com o uso da estrutura hierárquica, pode-se incluir todos os níveis de variáveis associadas, além de definir da melhor forma aquelas que explicam adequadamente a variável de interesse, ter ou não determinada doença.

Mostrou-se também, o impacto que o viés traz para a regressão logística calculando estimações diferentes e intervalos com grandes amplitudes, sendo estes fatos relacionados à prevalência da doença para cada situação testada, tal que à medida que a prevalência da doença se torna cada vez mais alta, há uma maior disparidade entre as estimações do modelo logístico e os outros modelos aqui definidos. Como visto, o modelo de Poisson pode apresentar um erro, já que sua distribuição base é a Poisson ao invés da binomial. Por isso, para a estimação da RP de forma correta para todos os tipos de prevalência se dá preferência ao modelo log-binomial, em que assume a distribuição exata para os dados e uma função de ligação que lineariza a relação e possibilita as inferências necessárias.

A aplicação da teoria foi feito com a utilização do software SAS e R para as estimações dos resultados. Os mesmos contém estruturas em seus programas que estão capacitadas para as devidas estimações e modelagens como a "PROC GenMod" e a função " GLM ". O SAS não conseguiu uma convergência direta para o modelo log-binomial, já o software R, por meio de um algoritimo que lida melhor com os limites que devem ser inseridos para as estimações dos parâmetros convergiu e definiu as estimações para a RP pela modelagem. Além disso, o software R, por ser gratuito e funcionar de forma colaborativa, contém outros pacotes que possuem opções de algoritmos que lidam com essa convergência, são eles, o "lbreg" e o "logbin". O código usado também está disponibilizado em um anexo neste trabalho.

# 5 Referências Bibliográficas

AGRESTI A. **An Introduction to Categorical Data Analysis**. 2. Ed. Florida, Gainesville, 2007.

ANANIAS M. C. **Implementação Computacional do Modelo Log Binomial.** 2015. 33 f. Trabalho de conclusão de curso (Curso superior em estatística) - Universidade de Brasília, Brasília, 2015.

BARRETO, S. M.; LIMA-COSTA, M. F. Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento. **Epidemiol. Serv. Saúde**, Brasília, v. 12, n. 4, p. 189-201, 2003.

BASTOS G. A. N.; HARZHEIM E.; SOUZA A. I. Prevalência e fatores associados à consulta médica entre adultos de uma comunidade de baixa renda do Sul do Brasil. **Epidemiol. Serv. Saúde**, Brasília, v.23 n.3, p.409 - 420, 2014.

BRAGANÇA, L. S. Introdução a análise estatística de variáveis dicotômicas e aplicações em dados socioeconômicos. **Revista brasileira de economia**, Rio de Janeiro, v.31, n.2, p. 323-352, 1977.

CABRAL I. S. C. **Aplicação do Modelo de Regressão Logística num Estudo de Mercado**. 2013. 59 f. Dissertção (Mestrado em Matemática Aplicada à Economia e à Gestão). Universidade de Lisboa. Lisboa, 2013.

COUTINHO L. M. S.; SCAZUFCA M.; MENEZES P. R. Métodos para estimar razão de prevalência em estudos de corte transversal. **Revista Saúde Pública**, São Paulo, v.42, n.6, p. 992-998, 2008.

DEDDENS J. A.; PETERSEN M. R. A comparison of two methods for estimating prevalence ratios. **BMC Medical Research Methodology**.Ohio, USA, 2008; DOI: 10.1186/1471-2288- 8-9.

ELIASZIW M.; WILIAMSON T.; FICK G. H. Log-binomial models: exploring failed convergence. **Emerging Themes in Epidemiology**, Canada, 2013.

GALLETTI A. J. F. **Comparação de métodos de estimação do risco relativo em estudos epidemiológicos.** 2015. 51 f. Trabalho de conclusão de curso (Curso superior em estatística) - Universidade de Brasília, Brasília, 2015.

HOSMER D. W.; LEMESHOW S.**Applied logistic regression**. 2. Ed. Canadá: John Wiley & Sons INC, 2001.

KUTNER M. H. et al. **Applied linear statistical models**. 5. Ed. New York: McGraw-Hill, 2004.

MARTINEZ E. Z.; OLIVEIRA M. F. S.; ROCHA J. S. Y. Fatores associados à cobertura vacinal em menores de cinco anos em Angola. **Revista Saúde Pública**, São Paulo, v.48, n.6, p. 906-915, 2014.

PAPELÉO C. L. M. **Estimação de risco relativo e razão de prevalência com desfecho binário.** 2009. 143 f. Dissertação (Programa de pós-graduação em epidemiologia) - Universidade federal do Rio Grande do Sul, Porto Alegre, 2009.

PAZÓ R. G. et al. Modelagem hierárquica de determinantes associados a internações por condições sensíveis à atenção primária no Espírito Santo, Brasil. **Cad. Saúde Pública**, Rio de janeiro, v.30 n.9, p.1891 - 1902, 2014.

TADANDO Y. S.; UGAYA C. M. L.; FRANCO A. T. Método de regressão de poisson:metodologia para avaliação do impacto da poluição atmosférica na saúde populacional. **Epidemiol. Ambiente Sociedade**, Campinas, v.7 n.2, p.405 - 414, 2019.

ZOU G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data.**American Journal of Epidemiology**. Canada, v.159, n.7, p. 702-706, 2004.

# 6 Código das programações utilizadas

Estão contidas abaixo as programações utilizadas em SAS e em R. As funções utilizadas são a PROC GenMod e GLM.

```
    Programação em SAS:
    proc import out=mysasdata
            datafile="/home/stuckert.alyne0/sasuser.v94/TCC/tspt.xlsx"
            dbms=xlsx replace;
   sheet='Plan1';
   run;
proc print;
run;

data a;set mysasdata;
run;

/* Variáveis Socio Econômicas */

proc format;
value escola 1-5 = 'Fundamental e Médio'
6-7 = 'Superior';
value rendaa 1 = 'Até 1 SM'
            2-4 = 'Acima de 1 SM';
run;

/* Variáveis Demográficas*/

proc format;
value idademae  14-30 = '14 a 30'
                31-44 = '31 a 44';
value idademaea 14-20 = '14-20'
                21-30 = '21-30'
```

```
31-44 = '31-44';
value civil 1 = 'Solteira'
            2-5= 'Não Solteira';
run;

/* Variáveis Durante Parto */

proc format;
value noparto 1 = 'Sim'
              2 = 'Não';
run;
/* Eventos Traumáticos */

proc format;
value trauma 1 = 'Sim'
             2 = 'Não';
run;

/* DPP */

proc format;
value dppa 1 = 'Com DPP'
           0 = 'Sem DPP';
run;

data b;set a;
if pontuedimburg1=. or IncendExplos=. or AcidGrave=. or AcidNaturais=. or
CrimeViolen=. or AbusoViolen=. or AbusoCrianca=. or
AlguemCrime=. or AmeacaVida=. or MorteInesp=. or EventViolent=. or
GraveNaofala=. or  ExperSexual=. or
MedoMorrer=. or Dorintensa=. or EquipInadeq=. or InformProced=. or
PerdaControl=. or ExperHumilh=. or
```

```
planejgesta=. or desejgesta=. or grupogesta=. or apoiopai=. or tipoparto=. or idade
pontutspt1=. then delete;
run;

data c;set b;
if pontuedimburg1 >= 11 then dpp = 1;
else dpp = 0;
run;

proc freq data=c;
table incendexplos;
run;

proc surveyfreq data=c;
tables (escolmae renda idademae estcivil)*dpp/cl row;
format escolmae escola. renda rendaa. idademae idademae. estcivil civil. dpp dppa.;
run;

proc surveyfreq data=c;
tables (IncendExplos AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave)*dpp/cl row;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil. dpp dppa.
IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.;
run;


proc surveyfreq data=c;
tables (MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh)*dpp/cl row;
```

```
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil. dpp dppa.
IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala ExperSexual
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh trauma.;
run;

/* Regressão logistica */

/* Bloco 1 */

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=escolmae /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Escolaridade Sup x Fund. e Médio' escolmae -1 1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=renda  /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Renda >1 x <1' renda 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=idademae  /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Idade 31-44 x 30-20' idademae -1  1/exp;
```

```
format escolmae escola. renda rendaa. idademae idademae. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp= estcivil /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Estcivil Nao Solt x Solt' estcivil 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=escolmae renda idademae estcivil /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Escolaridade Sup x Fund. e Médio' escolmae -1 1/exp;
estimate 'Renda >1 x <1' renda 1 -1/exp;
estimate 'Idade 31-44 x 14-30' idademae -1 1/exp;
estimate 'Estcivil Não Solt x Solt' estcivil 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademae. estcivil civil.;
run;


/* Bloco 2*/

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=IncendExplos /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
```

```
estimate 'IncendExplos Sim x Nao' IncendExplos -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AbusoViolen /dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AbusoViolen Sim x Nao' AbusoViolen -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AbusoCrianca /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AbusoCrianca Sim x Nao' AbusoCrianca -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
```

```
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AlguemCrime /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AlguemCrime Sim x Nao' AlguemCrime -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AmeacaVida /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AmeacaVida Sim x Nao' AmeacaVida -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
```

```
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=MorteInesp /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MorteInesp Sim x Nao' MorteInesp -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=EventViolent /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'EventViolent Sim x Nao' EventViolent -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
```

```
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=GraveNaofala /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'GraveNaofala Sim x Nao' GraveNaofala -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=ExperSexual /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'ExperSexual Sim x Nao' ExperSexual -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;
```

```
Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AcidGrave /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AcidGrave Sim x Nao' AcidGrave -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq idademae IncendExplos CrimeViolen AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=idademae IncendExplos CrimeViolen AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'IncendExplos Sim x Nao' IncendExplos -1 1/exp;
estimate 'CrimeViolen Sim x Nao' CrimeViolen -1 1/exp;
estimate 'AbusoViolen Sim x Nao' AbusoViolen -1 1/exp;
estimate 'AbusoCrianca Sim x Nao' AbusoCrianca -1 1/exp;
estimate 'AlguemCrime Sim x Nao' AlguemCrime -1 1/exp;
estimate 'AmeacaVida Sim x Nao' AmeacaVida -1 1/exp;
estimate 'MorteInesp Sim x Nao' MorteInesp -1 1/exp;
estimate 'EventViolent Sim x Nao' EventViolent -1 1/exp;
estimate 'GraveNaofala Sim x Nao' GraveNaofala -1 1/exp;
```

```
estimate 'ExperSexual Sim x Nao' ExperSexual -1 1/exp;
estimate 'AcidGrave Sim x Nao' AcidGrave -1 1/exp;format IncendExplos
AcidGrave AcidNaturais CrimeViolen AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademae. estcivil civil. escolmae escola.;
run;
run;

/* Bloco 3 */


Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=MedoMorrer
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MedoMorrer Sim x Nao' MedoMorrer -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=Dorintensa
```

```
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Dorintensa Sim x Nao' Dorintensa -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=EquipInadeq
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'EquipInadeq Sim x Nao' EquipInadeq -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=InformProced
```

```
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'InformProced Sim x Nao' InformProced -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=PerdaControl
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'PerdaControl Sim x Nao' PerdaControl -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=ExperHumilh
```

```
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'ExperHumilh Sim x Nao' ExperHumilh -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh
/dist=binomial link=logit type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MedoMorrer Sim x Nao' MedoMorrer -1 1/exp;
estimate 'Dorintensa Sim x Nao' Dorintensa -1 1/exp;
estimate 'EquipInadeq Sim x Nao' EquipInadeq -1 1/exp;
estimate 'InformProced Sim x Nao' InformProced -1 1/exp;
estimate 'PerdaControl Sim x Nao' PerdaControl -1 1/exp;
estimate 'ExperHumilh Sim x Nao' ExperHumilh -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
```

```
run;



/* Regressão de poisson */
/* Bloco 1 */

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=escolmae /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Escolaridade Sup x Fund. e Médio' escolmae -1 1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=renda  /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Renda >1 x <1' renda 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=idademae  /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Idade 31-44 x 30-20' idademae -1  1/exp;
format escolmae escola. renda rendaa. idademae idademae. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
```

```
model dpp= estcivil /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Estcivil Nao Solt x Solt' estcivil 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademaea. estcivil civil.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda idademae estcivil;
model dpp=escolmae renda idademae estcivil /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Escolaridade Sup x Fund. e Medio' escolmae -1 1/exp;
estimate 'Renda >1 x <1' renda 1 -1/exp;
estimate 'Idade 31-44 x 14-30' idademae -1 1/exp;
estimate 'Estcivil Nao Solt x Solt' estcivil 1 -1/exp;
format escolmae escola. renda rendaa. idademae idademae. estcivil civil.;
run;


/* Bloco 2*/

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=IncendExplos /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'IncendExplos Sim x Nao' IncendExplos -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
```

```
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AbusoViolen /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AbusoViolen Sim x Nao' AbusoViolen -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AbusoCrianca /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AbusoCrianca Sim x Nao' AbusoCrianca -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;
```

```
Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AlguemCrime /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AlguemCrime Sim x Nao' AlguemCrime -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=AmeacaVida /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AmeacaVida Sim x Nao' AmeacaVida -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
```

```
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=MorteInesp /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MorteInesp Sim x Nao' MorteInesp -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=EventViolent /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'EventViolent Sim x Nao' EventViolent -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
```

```
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=GraveNaofala /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'GraveNaofala Sim x Nao' GraveNaofala -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=ExperSexual /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'ExperSexual Sim x Nao' ExperSexual -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq escolmae renda IncendExplos /*CrimeViolen*/ AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
```

```
model dpp=AcidGrave /dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'AcidGrave Sim x Nao' AcidGrave -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
idademae idademaea. estcivil civil. escolmae escola. renda rendaa.;
run;

Proc GenMod data=c descending;
class NumPesq idademae IncendExplos AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave;
model dpp=idademae IncendExplos AbusoViolen AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual AcidGrave/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'IncendExplos Sim x Nao' IncendExplos -1 1/exp;
estimate 'AbusoViolen Sim x Nao' AbusoViolen -1 1/exp;
estimate 'AbusoCrianca Sim x Nao' AbusoCrianca -1 1/exp;
estimate 'AlguemCrime Sim x Nao' AlguemCrime -1 1/exp;
estimate 'AmeacaVida Sim x Nao' AmeacaVida -1 1/exp;
estimate 'MorteInesp Sim x Nao' MorteInesp -1 1/exp;
estimate 'EventViolent Sim x Nao' EventViolent -1 1/exp;
estimate 'GraveNaofala Sim x Nao' GraveNaofala -1 1/exp;
estimate 'ExperSexual Sim x Nao' ExperSexual -1 1/exp;
estimate 'AcidGrave Sim x Nao' AcidGrave -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
```

```
idademae idademae. estcivil civil. escolmae escola.;
run;


/* Bloco 3 */


Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=MedoMorrer
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MedoMorrer Sim x Nao' MedoMorrer -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=Dorintensa
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'Dorintensa Sim x Nao' Dorintensa -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
```

```
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=EquipInadeq
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'EquipInadeq Sim x Nao' EquipInadeq -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=InformProced
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'InformProced Sim x Nao' InformProced -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
```

```
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=PerdaControl
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'PerdaControl Sim x Nao' PerdaControl -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=ExperHumilh
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'ExperHumilh Sim x Nao' ExperHumilh -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
```

```
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;

Proc GenMod data=c descending;
class NumPesq idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh;
model dpp=idademae MorteInesp MedoMorrer Dorintensa EquipInadeq
InformProced PerdaControl ExperHumilh
/dist=poisson link=log type3 wald;
repeated subject=NumPesq/type=unstr;
estimate 'MedoMorrer Sim x Nao' MedoMorrer -1 1/exp;
estimate 'Dorintensa Sim x Nao' Dorintensa -1 1/exp;
estimate 'EquipInadeq Sim x Nao' EquipInadeq -1 1/exp;
estimate 'InformProced Sim x Nao' InformProced -1 1/exp;
estimate 'PerdaControl Sim x Nao' PerdaControl -1 1/exp;
estimate 'ExperHumilh Sim x Nao' ExperHumilh -1 1/exp;
format IncendExplos AcidGrave AcidNaturais CrimeViolen AbusoViolen
AbusoCrianca
AlguemCrime AmeacaVida MorteInesp EventViolent GraveNaofala
ExperSexual trauma.
MedoMorrer Dorintensa EquipInadeq InformProced PerdaControl
ExperHumilh noparto.
idademae idademae. estcivil civil. escolmae escola.;
run;
run;
```

Código em R:

```
getwd()
```

```
setwd("C:\\Users\\stuck\\Desktop\\bkp\\BKP\\Alyne\\documentos\\unb\\TCC\\arquivos base_r
dados <- read.csv2("tspt.csv")
dados2 <- read.csv2("tspt2.csv")
require(lbreg)

#DPP
summary(dados\$PontuEdimburg1)
dados\$PontuEdimburg1

cut(dados\$PontuEdimburg1, breaks=c(-1,10,30))
dados\$dpp<-cut(dados\$PontuEdimburg1, breaks=c(-1,10,30),labels = c('0','1'))
dados\$dpp

dados=subset(dados,is.na(dpp)!=T )


dados\$dpp=as.numeric(as.character(dados\$dpp))



#BLOCO 1

#ESCOLARIDADE
summary(dados\$EscolMae)
cut(dados\$EscolMae, breaks=c(-1,5,7))
dados\$escmae<-cut(dados\$EscolMae, breaks=c(-1,5,7),labels = c('EM','ES'))
dados\$escmae

dados=subset(dados,is.na(escmae)!=T)
es<-lbreg(dados\$dpp~dados\$escmae)
es
summary(es)
relrisk(es)
```

```
#glm
es2 <- glm(dpp ~ escmae, data = dados, family = binomial(link = log))
es2
summary(es2)
exp(coef(es2))


#RENDA

summary(dados\$Renda)
cut(dados\$Renda, breaks=c(0,1,4))
dados\$renda<-cut(dados\$Renda, breaks=c(0,1,4),labels = c('1SM','MSM'))
dados\$renda

dados=subset(dados,is.na(renda)!=T)
rm<-lbreg(dados\$dpp~dados\$renda)
rm
summary(rm)
relrisk(rm)

#glm
rm2 <- glm(dpp ~ renda, data = dados, family = binomial(link = log))
rm2
summary(rm2)

#IDADE

summary(dados\$IdadeMae)
cut(dados\$IdadeMae, breaks=c(14,30,44))
dados\$idadem<-cut(dados\$IdadeMae, breaks=c(14,30,44),labels = c('14a30','31a44'))
dados\$idadem
```

```
dados=subset(dados,is.na(idadem)!=T)
idm<-lbreg(dados\$dpp~dados\$idadem)
idm
summary(idm)
relrisk(idm)

#glm
idm2 <- glm(dpp ~ idadem, data = dados, family = binomial(link = log))
idm2
summary(idm2)
exp(coef(idm2))
li <- exp(0.5381 - (1.96*0.2246))
ls <- exp(0.5381 + (1.96*0.2246))

#ESTADO CIVIL
summary(dados\$EstCivil)
cut(dados\$EstCivil, breaks=c(0,1,5))
dados\$estcivil<-cut(dados\$EstCivil, breaks=c(0,1,5),labels = c('solteira','nsolteira'))
dados\$estcivil

dados=subset(dados,is.na(estcivil)!=T)
ec<-lbreg(dados\$dpp~dados\$estcivil)
ec
summary(ec)
relrisk(ec)

#glm
ec2 <- glm(dpp ~ estcivil, data = dados, family = binomial(link = log))
ec2
summary(ec2)
exp(coef(ec2))
```

```
li <- exp(0.02632  - (1.96*0.25100))
ls <- exp(0.02632 + (1.96*0.25100))

#MODELO BLOCO 1 COMPLETO
completo<-lbreg(dados\$dpp~dados\$escmae + dados\$renda + dados\$idadem + dados\$es
completo
summary(completo)
relrisk(completo)

#glm

completo <- glm(dpp ~ escmae + renda + idadem + estcivil,
                data= dados, family = binomial(link = log))

completo
summary(completo)
exp(coef(completo))
#escmae 0,7815
li <- exp(0.07114  - (1.96*0.25642))
ls <- exp(0.07114 + (1.96*0.25642))

#renda 0,2905
li <- exp(-0.27018  - (1.96*0.25558))
ls <- exp(-0.27018 + (1.96*0.25558))

#idade 0,0117
li <- exp(0.57612  - (1.96*0.22856))
ls <- exp(0.57612 + (1.96*0.22856))

#estado civil 0,9956
li <- exp(0.57612  - (1.96*0.22856))
ls <- exp(0.57612 + (1.96*0.22856))
```

```
#BLOCO 2

#incendio/explosão
summary(dados\$IncendExplos)
dados\$IncendExplos <- ifelse(dados\$IncendExplos == 2, 0, 1)
cut(dados\$IncendExplos, breaks=c(-1,0,2))
dados\$IncendExplos<-cut(dados\$IncendExplos, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$IncendExplos

dados=subset(dados,is.na(IncendExplos)!=T)

ie2 <- glm(dpp ~ IncendExplos, data= dados, family = binomial(link = log))
ie2
summary(ie2)
exp(coef(ie2))
li <- exp( 0.01224  - (1.96* 0.52488 ))
ls <- exp(0.01224 + (1.96* 0.52488 ))


# abuso/violência
summary(dados\$AbusoViolen)
dados\$AbusoViolen <- ifelse(dados\$AbusoViolen == 2, 0, 1)
cut(dados\$AbusoViolen, breaks=c(-1,0,2))
dados\$AbusoViolen<-cut(dados\$AbusoViolen, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$AbusoViolen
dados=subset(dados,is.na(AbusoViolen)!=T)
#glm
av2 <- glm(dpp ~ AbusoViolen, data= dados, family = binomial(link = log))
av2
summary(av2)
exp(coef(av2))
```

```
li <- exp( 0.8256   - (1.96* 0.2756))
ls <- exp(0.8256 + (1.96* 0.2756))


#abuso criança
summary(dados\$AbusoCrianca)
dados\$AbusoCrianca <- ifelse(dados\$AbusoCrianca == 2, 0, 1)
cut(dados\$AbusoCrianca, breaks=c(-1,0,2))
dados\$AbusoCrianca<-cut(dados\$AbusoCrianca, breaks=c(-1,0,2),labels = c('não','si
dados\$AbusoCrianca
dados=subset(dados,is.na(AbusoCrianca)!=T)
#glm
ac2 <- glm(dpp ~ AbusoCrianca, data= dados, family = binomial(link = log))
ac2
summary(ac2)
exp(coef(ac2))
li <- exp( 0.6652   - (1.96* 0.2769))
ls <- exp( 0.6652 + (1.96* 0.2769))

#Crime/lesão grave
summary(dados\$AlguemCrime)
dados\$AlguemCrime <- ifelse(dados\$AlguemCrime == 2, 0, 1)
cut(dados\$AlguemCrime, breaks=c(-1,0,2))
dados\$AlguemCrime<-cut(dados\$AlguemCrime, breaks=c(-1,0,2),labels = c('não','sim'
dados\$AlguemCrime
dados=subset(dados,is.na(AlguemCrime)!=T)
#glm

acr2 <- glm(dpp ~ AlguemCrime, data= dados, family = binomial(link = log))
acr2
summary(acr2)
exp(coef(acr2))
```

```
li <- exp(  0.4733   - (1.96* 0.2467))
ls <- exp(  0.4733 + (1.96* 0.2467))

#ameaça a vida
summary(dados\$AmeacaVida)
dados\$AmeacaVida <- ifelse(dados\$AmeacaVida == 2, 0, 1)
cut(dados\$AmeacaVida, breaks=c(-1,0,2))
dados\$AmeacaVida<-cut(dados\$AmeacaVida, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$AmeacaVida
dados=subset(dados,is.na(AmeacaVida)!=T)
#glm

av2 <- glm(dpp ~ AmeacaVida, data= dados, family = binomial(link = log))
av2
summary(av2)
exp(coef(av2))
li <- exp( 0.4726   - (1.96* 0.2703))
ls <- exp(0.4726 + (1.96* 0.2703))

#Morte inesperada

summary(dados\$MorteInesp)
dados\$MorteInesp <- ifelse(dados\$MorteInesp == 2, 0, 1)
cut(dados\$MorteInesp, breaks=c(-1,0,2))
dados\$MorteInesp<-cut(dados\$MorteInesp, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$MorteInesp
dados=subset(dados,is.na(MorteInesp)!=T)
#glm
mi2 <- glm(dpp ~ MorteInesp, data= dados, family = binomial(link = log))
mi2
summary(mi2)
exp(coef(mi2))
```

```
li <- exp( 0.5681   - (1.96* 0.2236))
ls <- exp(0.5681 + (1.96* 0.2236))
#evento violento

summary(dados\$EventViolent)
dados\$EventViolent <- ifelse(dados\$EventViolent == 2, 0, 1)
cut(dados\$EventViolent, breaks=c(-1,0,2))
dados\$EventViolent<-cut(dados\$EventViolent, breaks=c(-1,0,2),labels = c('não','si
dados\$EventViolent
#glm
ev2 <- glm(dpp ~ EventViolent, data= dados, family = binomial(link = log))
ev2
summary(ev2)
exp(coef(ev2))
li <- exp( -0.02068   - (1.96* 0.41490))
ls <- exp(-0.02068 + (1.96* 0.41490))
#evento grave e não fala

summary(dados\$GraveNaofala)
dados\$GraveNaofala <- ifelse(dados\$GraveNaofala == 2, 0, 1)
cut(dados\$GraveNaofala, breaks=c(-1,0,2))
dados\$GraveNaofala<-cut(dados\$GraveNaofala, breaks=c(-1,0,2),labels = c('não','si
dados\$GraveNaofala
dados=subset(dados,is.na(GraveNaofala)!=T)
#glm
gf2 <- glm(dpp ~ GraveNaofala, data= dados, family = binomial(link = log))
gf2
summary(gf2)
exp(coef(gf2))
li <- exp( 0.6577    - (1.96* 0.3191))
ls <- exp(0.6577  + (1.96* 0.3191))
```

```
#experiencia sexual violenta

summary(dados\$ExperSexual)
dados\$ExperSexual <- ifelse(dados\$ExperSexual == 2, 0, 1)
cut(dados\$ExperSexual, breaks=c(-1,0,2))
dados\$ExperSexual<-cut(dados\$ExperSexual, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$ExperSexual
#glm
es2 <- glm(dpp~ExperSexual,data = dados,family = binomial(link = log))
es2
summary(es2)
exp(coef(es2))
li <- exp(0.6577   - (1.96* 0.3191))
ls <- exp(0.6577 + (1.96* 0.3191))


#acidente grave

summary(dados\$AcidGrave)
dados\$AcidGrave <- ifelse(dados\$AcidGrave == 2, 0, 1)
cut(dados\$AcidGrave, breaks=c(-1,0,2))
dados\$AcidGrave<-cut(dados\$AcidGrave, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$AcidGrave
#glm
ag2 <- glm(dpp~AcidGrave,data = dados,family = binomial(link = "log"))
ag2
summary(ag2)
exp(coef(ag2))
li <- exp( -0.5449   - (1.96* 0.9423))
ls <- exp( -0.5449 + (1.96* 0.9423))
```

```
#bloco 2 completo
#glm mudando o start
completo3 <- glm(dpp ~idadem +IncendExplos + AbusoViolen + AbusoCrianca +
                   AlguemCrime + AmeacaVida + MorteInesp + EventViolent +
                   GraveNaofala + ExperSexual + AcidGrave, data=dados, family =
                   binomial(link = "log"), start = c(-1,0,0,0,0,0,0,0,0,0,0,0) )
completo3
summary(completo3)
exp(coef(completo3))

#incendio 0,8792
li <- exp(-0.08535  - (1.96*0.56173))
ls <- exp(-0.08535 + (1.96*0.56173))

#abuso violencia 0,7083
li <- exp(0.14408  - (1.96*0.38515))
ls <- exp(0.14408 + (1.96*0.38515))

#abuso criança 0,2319
li <- exp(0.41572  - (1.96*0.34771))
ls <- exp(0.41572 + (1.96*0.34771))

#algum crime 0,2319
li <- exp(0.27981  - (1.96*0.23501))
ls <- exp(0.27981 + (1.96*0.23501))

#ameça vida 0,7603
li <- exp( 0.09595  - (1.96*0.31454))
ls <- exp( 0.09595 + (1.96*0.31454))

#morte inesperada 0,0157
li <- exp(0.55698  - (1.96*0.23216))
```

```
ls <- exp(0.55698 + (1.96*0.23216))

#evento violento 0,3534
li <- exp(-0.36183  - (1.96*0.38988))
ls <- exp(-0.36183 + (1.96*0.38988))

#grave nao fala 0,1513
li <- exp(0.46566  - (1.96*0.32451))
ls <- exp(0.46566 + (1.96*0.32451))

#experiencia sexual 0,3972
li <- exp( 0.24174  - (1.96*0.28555))
ls <- exp( 0.24174 + (1.96*0.28555))

#acidente grave 0,3654
li <- exp(-0.79129  - (1.96*0.87427))
ls <- exp( -0.79129 + (1.96*0.87427))
#bloco 3

#medo de morrer
summary(dados\$MedoMorrer)
dados\$MedoMorrer <- ifelse(dados\$MedoMorrer == 2, 0, 1)
cut(dados\$MedoMorrer, breaks=c(-1,0,2))
dados\$MedoMorrer<-cut(dados\$MedoMorrer, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$MedoMorrer
dados=subset(dados,is.na(MedoMorrer)!=T )
#glm

mm2 <- glm(dpp~MedoMorrer,data = dados,family = binomial(link = "log"))
mm2
summary(mm2)
exp(coef(mm2))
```

```
li <- exp(0.9717   - (1.96* 0.2602))
ls <- exp(0.9717 + (1.96* 0.2602))

#dor intensa
summary(dados\$Dorintensa)
dados\$Dorintensa <- ifelse(dados\$Dorintensa == 2, 0, 1)
cut(dados\$Dorintensa, breaks=c(-1,0,2))
dados\$Dorintensa<-cut(dados\$Dorintensa, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$Dorintensa
dados=subset(dados,is.na(Dorintensa)!=T )
#glm

di2 <- glm(dpp~Dorintensa,data = dados,family = binomial(link = "log"))
di2
summary(di2)
exp(coef(di2))
li <- exp(  0.3197   - (1.96* 0.2574))
ls <- exp(  0.3197 + (1.96* 0.2574))

#equipe inadequada
summary(dados\$EquipInadeq)
dados\$EquipInadeq <- ifelse(dados\$EquipInadeq == 2, 0, 1)
cut(dados\$EquipInadeq, breaks=c(-1,0,2))
dados\$EquipInadeq<-cut(dados\$EquipInadeq, breaks=c(-1,0,2),labels = c('não','sim'
dados\$EquipInadeq
dados=subset(dados,is.na(EquipInadeq)!=T )
#glm

ei2 <- glm(dpp~EquipInadeq,data = dados,family = binomial(link = "log"))
ei2
summary(ei2)
exp(coef(ei2))
```

```
li <- exp( -0.2680   - (1.96*  0.4250))
ls <- exp( -0.2800 + (1.96*  0.4250))

#informação dos procedimentos
summary(dados\$InformProced)
dados\$InformProced<- ifelse(dados\$InformProced == 2, 0, 1)
cut(dados\$InformProced, breaks=c(-1,0,2))
dados\$InformProced<-cut(dados\$InformProced, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$InformProced
dados=subset(dados,is.na(InformProced)!=T )
#glm

ip2 <- glm(dpp˜InformProced,data = dados,family = binomial(link = "log"))
ip2
summary(ip2)
exp(coef(ip2))
li <- exp(  -0.4855   - (1.96*  0.2717))
ls <- exp(  -0.4855 + (1.96*  0.2717))

#perda de controle
summary(dados\$PerdaControl)
dados\$PerdaControl<- ifelse(dados\$PerdaControl == 2, 0, 1)
cut(dados\$PerdaControl, breaks=c(-1,0,2))
dados\$PerdaControl<-cut(dados\$PerdaControl, breaks=c(-1,0,2),labels = c('não','sim'))
dados\$PerdaControl
dados=subset(dados,is.na(PerdaControl)!=T )
#glm

pc2 <- glm(dpp˜PerdaControl,data = dados,family = binomial(link = "log"))
pc2
summary(pc2)
exp(coef(pc2))
```

```
li <- exp(  0.6888   - (1.96*  0.2244))
ls <- exp(  0.6888 + (1.96*  0.2244))

#experiencia humilhante

summary(dados\$ExperHumilh)
dados\$ExperHumilh<- ifelse(dados\$ExperHumilh == 2, 0, 1)
cut(dados\$ExperHumilh, breaks=c(-1,0,2))
dados\$ExperHumilh<-cut(dados\$ExperHumilh, breaks=c(-1,0,2),labels = c('não','sim'
dados\$ExperHumilh
dados=subset(dados,is.na(ExperHumilh)!=T )
#glm

eh2 <- glm(dpp~ExperHumilh,data = dados,family = binomial(link = "log"))
eh2
summary(eh2)
exp(coef(eh2))
li <- exp(  1.1467   - (1.96*  0.2993))
ls <- exp(  1.1467 + (1.96*  0.2993))
#modelo completo
completobl3 <- glm(dpp~idadem + MorteInesp + MedoMorrer + Dorintensa + EquipInadeq
                   + InformProced + PerdaControl + ExperHumilh,
                   data = dados,family = binomial(link = "log"),
                   start = c(-1,0,0,0,0,0,0,0,0))
completobl3
summary(completobl3)
exp(coef( completobl3))

#medo de morrer 0,0075
li <- exp(  0.71667   - (1.96*   0.26801))
ls <- exp(  0.71667 + (1.96*  0.26801))
```

```
#dor intensa 0,4797
li <- exp(  0.18486   - (1.96*   0.26156))
ls <- exp(  0.18486 + (1.96* 0.26156))

#equipe inadequada 0,4051
li <- exp(  -0.28309   - (1.96*  0.34002))
ls <- exp(  -0.28309 + (1.96*  0.34002))
#informação procedente 0,8345
li <- exp(  -0.03139   - (1.96*   0.15027))
ls <- exp(  -0.03139 + (1.96*   0.15027))
#perda de controle 0,2777
li <- exp(  0.26884   - (1.96*   0.24768))
ls <- exp(  0.26884 + (1.96*  0.24768))
#experiencia humilhante 0,0171
li <- exp(  0.58481   - (1.96*   0.24521))
ls <- exp(  0.58481 + (1.96*  0.24521))
```